


# Digitális szövegfolklor-adatbázisok és a számítógépes folklorisztika elméleti, módszertani és technikai kérdései<sup>1</sup>

## Digitális bölcsészettudomány, digitális textológia és a folklorisztika<sup>2</sup>

Az elmúlt húsz évben a mindennapi életet szinte észrevétlenül formálta át a digitális technológia, ami a tudománytörténészek és filozófusok szerint jelentősebb fordulópont az emberi kultúrában, mint a nyomtatás feltalálása volt a reneszánszban (Dávidházi 2014; McGann 2014). Az új technológia a tudományos kutatás kereteit és lehetőségeit is gyökeresen megváltoztatta. Habár gyakorlati alkalmazása evidens – azaz ma már szinte elképzelhetetlen olyan kutatás, amely valamilyen módon ne használná a számítógépet vagy a digitális világ nyújtotta egyéb lehetőségeket, a bölcsészek főként a digitalizált tartalmakat –, egyelőre vitatott, hogy mindez csak szemléletváltást, új módszertant, vagy esetleg egy új, önálló diszciplínát eredményez a bölcsészettudományokon belül (Kokas 2016: 405; Thomas 2016).<sup>3</sup>

1 Az ezekhez az eredményekhez vezető kutatás az Európai Kutatási Tanács részéről, az Európai Közösség hetedik keretprogramjából (2007–2013), az EKT 324214 sz. támogatási megállapodása alapján finanszírozásban részesült; a tanulmány az  Emberi Erőforrások Minisztériuma ÚNKP-17-3 kódszámú Új Nemzeti Kiválóság Programjának támogatásával készült.

2 A tanulmány korábbi előadás-vázlatai a Folklore Fellows' Summer School 2015-ös turkui „Doing Folkloristics in the Digital Age” elnevezésű nyári egyetemén, a SIEF Working Group on Archives 2015-ös zágrábi kongresszusán, valamint a Kriza János Néprajzi Társaság nemzetközi vándorkonferenciáján („Az archívumok használatának változó kontextusai” 2016. október 14–15-e, Ottomány [Otomani], Románia) hangzottak el. A fent említett találkozókon kapott számos hozzászólásért és építő jellegű kritikáért ezúton mondok köszönetet. Köszönettel tartozom továbbá az MTA BTK Néprajztudományi Intézet „Kelet–Nyugat” Vallásnéptudományi Kutatócsoport valamennyi munkatársának, akik több alkalommal is véleményezték a szöveg korábbi változatait.

3 A digitális bölcsészettudomány (*digital humanities*) recens kézikönyvei, összefoglalásai: Schreibman, Siemens és Unsworth (2016); Berry (2012).

A szemünk előtt zajló folyamat tehát a tudományos forradalmak szerkezetének megfelelően történik, egymással versengő módszerek, elméletek és terminológiai zűrzavar jellemzi,<sup>4</sup> végkimenetele pedig egyelőre bizonytalan és ismeretlen.<sup>5</sup> Miközben a digitális technika a legtöbb humán tudományi diszciplínát alapvetően mozdította ki jól megalapozott pozíciójából (és éppen ezért a bölcsészársadalom egyik fele gyanakvással és idegenkedve tekint rá), az angolszász nyelvterületen *Digital Humanities* néven felemelkedett új tudományág kiépítette saját intézményrendszerét (nemzetközi társaságok, tanszékek és szakok alapítása, évenként megrendezett kongresszusok stb.),<sup>6</sup> nemzetközi historiográfusai pedig már a tudományszakon belüli harmadik hullámról/fordulatról beszélnek.<sup>7</sup>

A digitális fordulat(ok) elsődlegesen a bölcsészettudomány egyik legfontosabb területét, a filológiai-textológiai munkát, azaz a *szöveg-előállítás, szövegkiadás* alapevékenységét érintették, ami jól érzékelhető a digitális bölcsészethez köthető kiadványok és konferenciák gyors áttekintésével is, azok túlnyomó része ugyanis a *digitális textológia* kérdéseit és problémáit járja körül (McGann 2014; Dávidházi 2014; Debreczeni 2014; Schreibman, Siemens és Unsworth 2016). Ezzel mintegy párhuzamosan a digitális fordulat kedvezett annak a már korábban megkezdődött folyamatnak is, amely által a textológia kilépett sokáig lenézett pozíciójából – miszerint csupán a bölcsészettudomány segédtudománya, „szolgálóleánya” lenne (McGann 2014: 19–20; Szilágyi 2014) –, hiszen a különféle posztmodern szövegelméleteknek a statikus, analóg kritikai kiadás nem tudott megfelelni, a digitális médium dinamikus és rugalmas lehetőségeivel azonban arra alkalmas(abb) eszköznek bizonyul (Debreczeni 2014: 27–28).

A fenti átalakulások a folklorisztikát is a tudományszak lényegi, alapvető problémáihoz és kérdéseihöz vezetik vissza. Az adatgyűjtési, archiválási, textológiai eljárásokból és technikákból fakadó módszertani és elméleti dilemmák a folklorikatások kezdete óta jelen vannak, a róluk való élénk diskurzus pedig az elmúlt több mint százötven évben nem egy esetben éles tudománytörténeti fordulópontot hozott.<sup>8</sup> Éppen ezért a digitális bölcsészettudomány alakuló paradigmájában is szükségszerű az alábbiakat újra vita tárgyává tenni: (1) Hogyan érdemes tárolni és kiadni a folkloristák által és a kutatás során felhalmozott hatalmas meny-

---

4 'Computing humanities' és 'digital humanities', illetve a különféle 'computational' és 'digital' előtaggal ellátott diszciplínák. A magyar szóhasználatban a *digitális bölcsészettudomány(ok)* kezd meghonosodni, az SZTE, a DTE és a PPKE is ezzel a névvel indított digitális bölcsészet szakot, amelyet időközben sajnos megszüntettek. Néhány tanulmányban előfordul a *digitális humán tudomány* is, illetve a *bölcsész informatika*, mely korábban külön programként (BIÖP) működött az ELTE BTK-n. Ugyanakkor számítógépes nyelvészetet több helyen is lehet tanulni jelenleg az országban (pl.: ELTE BTK, DTE, PPKE, SZTE stb.) A folklorisztikát érintően is két terminus van használatban, a 'digital folkloristics' és a 'computational folkloristics'. Ehhez vö. Vargha (2016).

5 A digitális bölcsészettudomány intézményesülését tudománytörténeti szempontból Claire Warwick az anglistika 19. századi harcához hasonlította tanulmányában (Warwick 2016).

6 A digitális bölcsészettudomány intézményrendszeréhez lásd Thomas (2016).

7 Az első hullámot az 1990-es évek végén és a 2000-es évek elején főként kvalitatív megközelítés jellemezte, azaz minél több anyag digitalizálása, valamint e nagy szövegtörzsek elemzésére alkalmas infrastruktúra kialakítása volt a cél. A kutatók úgy tekintettek a digitális technológiára, mint egy kiegészítő, gyorsító eszközre. A szóhasználatot tekintve erre a periódusra a 'computing in the humanities' és 'humanities computing' volt a jellemző. A második hullámnál a névváltással együtt (*digital humanities*) azonban már önálló diszciplináris igények lépnek fel, azaz sajátos és hibrid módszerek és eszközök alkalmazása, olyan publikációs modellek, melyek már nem a könyvkultúrából származnak, és kiemelten foglalkoznak az eredendően digitális anyagok problematikájával (Berry 2011: 2–4, 2012).

8 A folklorisztika textualizációs paradigmáihoz lásd: Fine (1984); Foley (1997 [1995]); Voigt (2004); Honko (2000a); Landgraf (2006); Seitel (2012); Niles (2013a, 2013b); Katajamäki és Lukin (2013). Magyar folklortextológiai útmutató: Voigt és Balogh (1974); Barna (2003).

nyiségű szöveganyagot? (2) Milyen eszközök és eljárások szükségesek ehhez? (3) Milyen új módszerekre van szükségünk az elemzésekhez? (4) Milyen elemzésekre alkalmasak és milyen elemzésekre nem alkalmasak a digitális folklórszövegtörzsek?

Tanulmányomban az európai néprajztudomány/folklorisztika intézményesülésekor, a 19. század végén, 20. század első felében létrejött nemzeti folklórárchívumok és gyűjtemények digitalizálási problémáit és lehetőségeit összegzem.<sup>10</sup> Írásom elsődleges célja néhány már elkészült és elérhető, folklórszöveget tartalmazó és feldolgozó digitális adatbázis és az azokon számítógépes eljárásokkal végzett elemzések problémaorientált és kritikus áttekintése.<sup>11</sup> A kérdést a folklorisztikai archiválás és a digitális bölcsészettudományok tágabb kontextusában tárgyalom. Elsőként a folklóradatbázisok tartalmi koncepcióinak különféle tudománytörténeti okait, majd a digitális tartalom-előállítás gyakorlati, technikai és módszertani-elméleti oldalát mutatom be. Ezt követően a digitalizált folklórszövegek elemzéséhez használt új digitális eszközöket és eljárásokat veszem górcső alá, végezetül az adatbázisok és a számítógépes folklorisztika elméleti alapvetéseit összegzem. A fentieket megelőzően röviden kitérek az európai folklórárchívumokra, valamint a folklorisztika és az archiválás kapcsolatára, hiszen a legtöbb digitális folklóradatbázis egy korábban létrejött folklórárchívum vagy archívumi anyag egy részének vagy több archívum anyagának együttes digitalizálására vállalkozik.

## Archívumok és archiválás a folklorisztikában

A különféle történeti levéltárak, kéziratárak és archívumok között az európai folklórárchívumok sajátos keverék képződmények.<sup>12</sup> Heterogenitásuk érzékeltetéséhez csak néhány fontos tényezőre szeretnék utalni röviden.<sup>13</sup> Az első a folklór gyűjtésének eltérő motivációi,

9 A számítógépes folklorisztika feladatait és célkitűzéseit illetően lásd Timothy R. Tangherlini programadó tanulmányait: Abello, Broadwell és Tangherlini (2012); Tangherlini (2013a, 2016a).

10 Jelen szövegben egységesen folklórárchívumnak nevezem a nemzetenként és országonként változó megnevezésű (néprajzi, folklór, etnológiai, etnográfiai, tradicionális-hagyományos) folklórányagot tartalmazó archívumokat, adattárakat. Lásd ehhez a SIEF Working Group on Archives állásfoglalását: <https://www.siefhome.org/wg/arch/> (letöltve: 2017. június 20.). Ezzel azt is hangsúlyozom, hogy írásom alapvetően folklorisztikai megközelítésű, azaz a korábban a szóbeliségből lejegyzett kéziratok, gépelt szövegek digitalizálásához kíván hozzájárulni.

11 Ebből kifolyólag nem térek ki egyéb fájlformátumok tárolásának elméleti, technikai és etikai kérdéseire (hangfájl, videó, fénykép stb.), valamint kisebb, veszélyeztetett kultúrák szóbeliségének megőrzési és örökségésítési problémájára – ahol nagyjából egy időben zajlik az adatgyűjtés és a gyűjtött anyag digitális archiválása. Ezenkívül csak olyan tudományos adatbázisokat veszek számba, melyek számítógépes programok által olvasható és ezáltal értelmezhető és elemezhető textualizációs eljárásokat követnek, ennélfogva tudományos kutatásra alkalmas eszközök.

12 A francia forradalomtól datálható modern kori különféle állami, kormányzati, történeti, intézeti publikus archívumok nemzetközi viszonyításban szintén nagyon eltérnek egymástól (Markoff 2015).

13 Az európai folklórárchívumok történetéről nem készült mind ez ideig áttekintő, összefoglaló munka, így csak néhány, a digitalizálás szempontjából fontosabb problémára térek ki. A felsoroltak nem általánosíthatók minden folklórárchívumra, nem azonos arányban, súlyal és időben jelentkeznek az egyes konkrét nemzeti esetekben. A Regina Bendix és Galit Hasan-Rokem által szerkesztett *A Companion to Folklore* című kézikönyv Bjarne Rogannak az európai folklorisztika intézményeiről írt fejezetével (északi, ír, holland, francia, német archívumok említésével) némiképp pótolja ezt a hiányt (Rogan 2012: 610–614). Andy Kolovos pedig az amerikai folklórárchívumokról készített disszertációt, melyben röviden utal azok európai hátterére, valamint az európai és amerikai folklórárchívumok főbb különbségeire (Kolovos 2010: 1–87). Az észak-európai folklórárchívumokról (Dánia, Feröer szigetek, Svédország, Norvégia, Izland, Finnország) fontos, katalógusszerű áttekintés készült 1978-ban (Herranen és Saressalo 1978).

valamint az azok mögött meghúzódó ideológiák változatossága (felvilágosodás, népnevelés, nemzetépítés, nacionalizmus, megörökítés, örökségésítés, hagyományörzés, tudományos kutatás).<sup>14</sup> Leválasztva e különféle indítatásokról a szűkebb értelemben vett – önmagában is folyamatosan változó – tudományos érdeklődés által létrehozott folklórgyűjteményeket,<sup>15</sup> még mindig sokféle és szerteágazó anyagot kapunk az egymás mellett párhuzamosan létező folklórdefiníciók és folklórgyűjtési technikák pluralitása miatt.<sup>16</sup> A tudományosan elismert, nagy múltú intézmények esetében a diszciplína átalakuló – és általában táguló – folklórmeghatározásával és a folklórananyag megörökítésére használt különféle módszerekkel ugyanis jelentősen változott a folklórként azonosított, folklórchívumokba bekerülő anyag és annak jellege is.<sup>17</sup> A folklórgyűjtemények tehát speciális képződmények abból a szempontból, hogy többségük meghatározott kutatási projekt mellékterméke, ennél fogva elsődlegesen befolyásolja az a tényező, hogy a kutatást végző személy/csoport kutatásának tárgyát és célját hogyan határozta meg, ennek elérésére pedig milyen módszer(eke)t választott (Wolf-Knuts 2001: 9–14). Ebből következik tehát az is, hogy a forrásdokumentumok típusai nem egyneműek. A gyűjtők vagy a kutatási programok hagyatékai sokféle anyagból állnak össze: cédulák, kérdőívek, jegyzetek, memoárok, terepnaplók, a terepmunka egyéb járulékos elemei, különféle magán- vagy hivatalos levelezések, pizskozatok, kiadásra előkészített tiszta-  
tázatok stb. Érdemes arra is felhívni a figyelmet, hogy az európai folklórchívumok meglehetősen különböző intézményi háttérrel és rendszerben működnek, s ez jelentős mértékben

---

14 A különböző ideológiai indítatáshoz és folklórgyűjtési koncepciókhoz a nemzetközi szakirodalmat illetően lásd Anttonen (2005); Wolf-Knuts (2001, 2010); Valk (2005); Baycroft és Hopkin (2012); Kuutma (2015).

15 Bár az *amatőr-professionális* gyűjtő szétválasztása a folklórchívumok anyagát tekintve valójában a legtöbb esetben nem túl üdvöztető és nem is lehetséges, mivel azok anyagának döntő többsége a 19. századtól kezdve lelkes önkéntes vagy fizetett amatőr gyűjtők szorgos munkájának eredménye. Forrai Ibolya kimutatása (1998-ig) szerint például a budapesti Néprajzi Múzeum Etnológiai Archívumának legnagyobb részét (több mint 12 000 tétel) a társadalmi gyűjtőhálózaton keresztül benyújtott pályamunkák alkotják (Forrai 2000: 33). Az önkéntes gyűjtőhálózatokat viszont nem feltétlenül követte mindig szigorú felső irányítás, egységes gyűjtési koncepció, vagy ha igen, akkor azok tényleges megvalósulásának a gyakorlatban nagy eltérések voltak. Éppen ezért már a kezdetektől a szakma professzionalista művelői jogosan félték a dilettantizmustól és próbálták a szakszerűség irányába terelni a gyűjtéseket. Erre hívta fel a figyelmet többek között a 19. század legvégén Katona Lajos, aki jelentős szerepet játszott a magyar folklorisztika tudománnyá válásában (Landgraf 2016: 509–510). Az ír teljes állású és részmunkaidős gyűjtők kontrolljához vö. Briody (2007: 415–429). Vagy ugyancsak ehhez vö. Kati Mikkola tanulmányát a finn folklórchívum önkéntes gyűjtőiről és konfliktusaikról a hivatásos folkloristákkal. Ez utóbbi ugyanakkor lehetőséget biztosít annak kutatására, hogy az önkéntes, autodidakta gyűjtők milyen indítatásból és milyen folklórfelfogás mentén gyűjtöttek (Mikkola 2013). A 19. századi laikus, önkéntes észti folklórgyűjtők tevékenységének vernakuláris írásbeliségként való értelmezéséhez vö. Kikas (2014).

16 Az egymás mellett élő pluralitásra számtalan példa hozható. A 19. század végi, 20. század eleji, éppen átalakuló textológiai paradigmában a magyar folklorisztika szaktekintélyei például sokféleképpen nyilatkoztak az emlékezetből gyűjtés, a diktáltatás, a gyorsírásos gyűjtés vagy a fonográf után történő lejegyzés különböző előnyeiről és hátrányairól, ki az egyiket, ki a másikat tartva hitelesebbnek (Gulyás 2015; Landgraf 2016: 511–513). Domokos Mariann arra hívja fel a figyelmet, hogy a gyűjtő fogalma is megváltozott a századforduló magyar folklorisztikájában, amikor is szerzőből pusztán dokumentáló személyekké váltak a gyűjtők (Domokos 2015). Ugyancsak vö. ehhez Fredrik Skott svéd tudománytörténeti kutatásait. Skott többek között arra mutatott rá, hogy a tudományos hatalomért és legitimációért az 1930-as évek Svédországában a svéd néprajzi atlasz előmunkálatai során a göteborgi Västsvenska Folkminnesarkivet (melynek akkori vezetője Carl-Martin Bergstrand) és az uppsalai, lundi és stockholmi folklórchívumok közötti parázs vitákból két teljesen ellentétes gyűjtési és archiválási módszertan és elképzelés bontható ki (Skott 2001, 2008).

17 A finn folklórchívum az 1960-as évektől kezdve inkább élettörténeteket, oral historyt gyűjtött kampányszerűen, különböző foglalkozásokra fókuszálva (pl.: kórházi dolgozók, útépítő munkások stb.) (Harvilahti 2012: 402–404). Romániai magyar vonatkozásban vö. a 20. század második felétől meginduló első ízű paraszti élettörténetek kutatását, később tanítók, mérnökök élettörténetének dokumentálását (Keszeg 2011: 165–194).

befolyásolja a gyűjtemények archiválási struktúráját és – ami a dolgozat szempontjából lényegesebb – ily módon annak digitalizálási lehetőségeit és kereteit is.<sup>18</sup> A folklórarchívumok hol egyetemi tanszékek részei, hol különféle kulturális, irodalmi, néprajzi és folklórtársaságok, szervezetek, kéziratárak alá tartoznak, máshol múzeumi (irodalmi, kulturális vagy kimondottan néprajzi) keretben működnek, ugyanakkor teljesen önálló intézményként is előfordulhatnak (Rogan 2012: 610). Végezetül fontos megemlíteni az archívumokat fenntartó intézmények aktuálpolitikai helyzetét is, azaz a gyűjtemény viszonyát a mindenkori állami vezetéshez, hiszen mindez szintén nagymértékben alakította és alakítja azok sorsát – a hatalomtól teljesen független tudomány képe illúzió.<sup>19</sup>

A fentiekben körvonalazott különbségek ellenére az európai folklórarchívumok mégis sok szállal kötődnek egymáshoz, legfőképpen azért, mert a néprajztudomány összehasonlító perspektívájából adódóan a kezdetektől fogva nemzetközi együttműködésben képzelte el önmagát, a komparatív törekvések fontos bázisát pedig az egyes nemzeti archívumok alkották, kiegészülve egy lehetséges európai folklórarchívummal, ami azonban soha nem valósult meg (Nic Craith 2008; Rogan 2012: 604–606, 2014: 174). A nemzetközi folklorisztika legnagyobb közös metszetét a diszciplínát sokáig meghatározó földrajztörténeti irányzatok és módszerek jelentették, amelyhez az európai néprajztudomány/folklorisztika nagy közös projektjei, mint például az *Enzyklopädie des Märchens*, a nemzetközi típuskatalógusok, a nemzeti és nemzetközi néprajzi és dialektológiai/nyelvi atlaszok köthetők (Schmitt 2005; Rogan 2014: 176–177). A többnyire ezek megvalósítására kiépített önkéntes gyűjtőhálózatok, a gyűjtési kérdőívek, útmutatók, az észak-európai, skandináv mintára kialakított katalógusrendszer miatt az európai archívumok törzsállományának jellege és azok struktúrája nagyfokú hasonlóságot mutat, amely a különbségek ellenére indokoltá teszi az archívumok digitalizálása során felmerülő elméleti, módszertani és technikai problémák együttes tárgyalását.<sup>20</sup>

Az archívumok a folklorisztika – és a tágan értelmezett néprajz – tudománnyá válásában kulcsfontosságú szerepet töltek be és meghatározó intézményei voltak a 20. század közepéig (Gulyás 2015: 18). Az 1960-as, 1970-es években bekövetkezett episztemológiai fordulatok azonban jelentősen átalakították a folklórarchívumok szerepét és értékét, főként az észak- és nyugat-európai, valamint az amerikai folklorisztikában, aminek következtében néhol teljes generációk tagadták meg e gyűjtemények használatát (Wolf-Knuts 2001:12; Kolovos 2010: 23; Gunnell et al. 2013: 173; Rogan 2012: 613–614; Anttonen 2013; Harvilahti 2012: 402–403).<sup>21</sup> Ahogy a posztmodern szövegelméleteknek nem felelt meg az

---

18 A fenntarthatóság szempontjából nagyon fontos, hogy az adatbázisok is valamilyen intézményi struktúra részei legyenek (Debreczeni 2014: 29), amihez állandó anyagi fedezet szükséges (Järv és Sarv 2014: 59). Ugyanakkor az adatbázis-építés jellegét és lehetőségét meghatározza, hogy az adott intézménynek milyen digitalizálási elképzelései és prioritásai vannak.

19 Például az *Irish Folklore Commission* archívumának megalapítása a független Írország megszületésekor (Briody 2007: 33–69). Szerencsére egyre többen kutatják a szovjet korszak és a folklórarchívumok viszonyát is (Västrik 2007; Kulasalu 2013).

20 A skandináv mintából indultak ki például a budapesti Néprajzi Múzeum Ethnológiai Adattárának létrehozói az 1930-as években (Forrai 2000: 614), az 1935-ben alapított Irish Folklore Commission archívuma pedig kimondottan az uppsalai rendszert vette át (Briody 2007: 325–331).

21 Persze e folyamatok megint eltérő erősséggel és ütemben jelentkeztek a különböző országok folklorisztikájában, de a nemzetközi kongresszusok és publikációk tematikájából Terry Gunnell szerint egyértelműen kirajzolódik egy ilyen tendencia (Gunnell et al. 2013: 172–173). Pontosan ezért érezte szükségét, hogy egy nemzetközi fórumon revideálja az archívumellenességet. A vilniusi ISFRN kongresszuson 2013-ban kerekasztal-beszélgetést szerveztek a témának, melynek teljes anyaga nyomtatásban is megjelent (Gunnell et al. 2013).

analóg könyv, ugyanúgy a folklorisztika antropológiai, pragmatikai fordulata után az archívum és annak merev struktúrája nem elégítette ki a diszciplína megváltozott igényeit.<sup>22</sup> A folklorisztika figyelme a múlttól a jelenre, a szövegről a performanszra és a használatra, a struktúráról és formáról a kontextusra és az interakcióra, a közösségről az egyénre, a falusíróról a városira, a szóbeliről az írásbelire (is) fordult, és ennek következtében az *archívumról* a *terepré*.<sup>23</sup> Ebből a perspektívából nézve az elődök által gyűjtött anyag vizsgálatát mellőzték azzal a – bizonyos szempontból helytálló, ám némiképp eltúlzott, sokak által viszont máig hangoztatott (Beyer 2011: 3) – gondolattal is megtámogatva, miszerint az archívumban lévő szöveganyag csupán annak kutatására alkalmas, hogy a folklorisztika egy adott periódusban mit tartott folklórnak (Gunnell et al. 2013: 171).

Bár a földrajztörténeti irányzat és a pozitivistá, kontextus nélküli adathalmaz elutasítása egyfelől a diszciplína termékeny megújulását eredményezte, módszereinek és elméleteinek teljes negligálása együtt járt az elődök által gyűjtött folklóranyag és a gyűjtési módszertan démonizálásával, ami az archívumokat több szempontból is válságos helyzetbe hozta. Új archiválási rendszert és struktúrát kellett (volna) kidolgozniuk, hiszen az új igényeknek az előző nem felelt meg; az időközben eltűnőben lévő hagyományos népi kultúra miatt pedig át kellett gondolniuk a folklórarchívumok gyűjtési körét és társadalmi szerepvállalását is. Míg az utóbbi két feladat viszonylag egyszerűbben kivitelezhető volt sok helyen, addig az intézményi rendszert nem lehet (és lehetett) egyik napról a másikra átalakítani, az archívumot pedig nem szabhatják a kutató(k) és a kutatás folyamatosan változó kérdéseihez. Az északi, skandináv országokban már az 1970-es évektől problémaként merült fel az is, hogy mi legyen a sorsa a korábbi rendszernek: szüntessék meg és kezdjenek egy újat, ezáltal párhuzamosan legyen kettő?<sup>24</sup> Nehezítette a kérdésekre adható válaszokat az is, hogy a legtöbb archívumban – főleg azokban az országokban, ahol csak a 20. század első harmadában jöttek létre e gyűjtemények – napi szinten az előző paradigma által felhalmozott gyűjtések rendszerezésének és katalogizálásának beláthatatlanságával kellett szembenézniük a kutatóknak.<sup>25</sup> Mindemellett érthető, hogy a társadalmi kontextus, a használat és a performansz felől értelmezett folklór (és azon belül is a szóbeliség) vizsgálatához az adatok részletesebb, a terepmunka kvalitatív tulajdonságait sokkal inkább megőrző, új dokumentálási eljárásra

---

22 Fredrik Skott a svéd folklórarchívumok kapcsán a vilniusi kerekasztal-beszélgetésen arra hívta fel a figyelmet, hogy az archívumi struktúra egyáltalán nem tudta követni a megváltozott folklórkoncepciókat, az archívumoktól való elfordulás legfőbb oka tulajdonképpen ebben rejlett (Gunnell et al. 2013: 199–200).

23 E folyamatokban természetesen közrejátszott további két tényező is. Az egyik, hogy a tudomány vizsgálati tárgya, az európai népi kultúra lassan mindenhol teljesen átalakult. A másik, hogy a paradigmátikus fordulatokat segítették a 20. század eleji technikai innovációk. A 20. században megjelent adatrögzítési technikák (videófelvétel, magnófelvétel) általánossá válása a folklorisztikában nagy szerepet játszott az egyéniségkutatás, a performatív és kontextualista irányzatok felívelésében (Voigt 1997; Kolovos 2004: 24). A folklorisztika pragmatikai fordulatához – azon belül is a performanszkutatásokról – összefoglalóan lásd Bauman (2012).

24 A felmerült kérdéseket lásd *NIF Newsletter* 1978 6(1); 1982 10(4); 1989 17(4).

25 A budapesti Néprajzi Múzeum Etnológiai Adattárának történetében tulajdonképpen csak a működés harmadik évtizedében, az 1960-as években tudtak nekilátni a felhalmozódott anyagok módszeres rendezésének. Az ehhez elengedhetetlenül szükséges véglegesített szakrend is csak 1967-ben jelent meg (Forrai 2000: 618–619). Seán Ó Súilleabháin, az ír folklór társaság archívumának akkori vezetője 1970-ben kissé szkeptikusan számol be a katalogizálómunka állásáról. Jelentésében arról ad hírt, hogy míg a gyűjtő, gyűjtési helyszín, adatközlő szerinti mutatók naprakészek, a beérkezett kéziratok tematikus mutatóinak elkészítéséhez nagyjából hat, teljes állásban csak katalogizáló munkatársra lenne szüksége az archívumnak, akik ezt a feladatot nagyjából 20 évnyi munkával tudnák elvégezni. Az írek esetében a katalogizálást nehezítette az a tényező is, hogy csak kevesen tudtak olyan szinten írni, hogy az ír nyelvű kéziratokat feldolgozzák. A katalogizálás menetéről és küzdelmeiről lásd részletesen Briody (2007: 325–331).



volt igénye a folkloristáknak. Az észak-európai és amerikai folklorisztikában pontosan azokat a kutatókat foglalkoztatták e kérdések, akik a legkritikusabban szólaltak fel az elődök ellen.<sup>26</sup> Lauri Honko írásaiban a folklorarchívumok korábbi anyagát kontextusa vesztett, „halott leletnek” minősítette,<sup>27</sup> ugyanakkor ezzel szinte egy időben a turkui egyetem munkatársaival olyan terepmunka-módszertan kidolgozását célozták meg, mely során az adatokat már eleve a későbbi archiválási rendszernek megfelelően rögzítik, a rögzítés módja pedig tartalmaz minden olyan információt a folkloranyagról, amely szerintük egy későbbi felhasználó számára elengedhetetlenül szükséges az értelmezéshez.<sup>28</sup>

A folklorisztikai adatrögzítés tudománytörténete szereti magát evolucionista fejlődésnarratívába ágyazni, mely szerint a technológia révén a gyűjtések egyre jobbak, az adatok egyre pontosabbak, a tudományos eredmények pedig ezáltal egyre hitelesebbek (Gulyás 2015: 24), azonban a szóbeliség megörökítésére bevezetett új technikák (magnófelvétel, videófelvétel), vagy az aprólékos, minden kis részletre kiterjedő dokumentáció valójában nem oldották meg a médiumváltás összes problémáját, az oralitás akadémiai kutatása pedig továbbra is olvasható szövegeket várt el és vár el a mai napig a kutatóktól (Honko 2000a: 30). Ugyan sokféle újító kísérlet született, mint például Charles Briggs könyve a mexikói szóbeliségről, ahol a szövegek két nyelven, már-már zenei partitúrához hasonlóan kerültek lejegyzésre a különféle paralingvisztikai elemek mellékjeles rögzítésével (gesztusok, intonáció, hangerő, mimika, kontextus, vö. Briggs [1988]), ám mindezek a kiadási gyakorlatot lényegesen nem formálták át. A performatív fordulat tehát csak megerősítette a kutatókat két fontos felismerésben. Egyrészt abban, hogy lehetetlen a torzítás nélküli médiumváltás, akármennyire tökéletes technikával és részletességgel rögzítjük is a szóbeli előadást; másrészt abban, hogy az archiválási és textológiai gyakorlatban nem lehet általános elveket lefektetni és nem lehet azokat teljes mértékben egységesíteni, mert mindig a kutatási célnak/kérdésnek leginkább megfelelő módszert kell alkalmazni (Honko 2000a: 29–36; Fine 1984; Finnegan 1992: 174–199).<sup>29</sup>

Az archívumok és a folklorisztikai textológia posztmodern kritikái ugyanakkor egy ponton szerencsésen lendítették fel a folklorarchívumi anyagok kutatását. Az elődök munkásságának kritikus vizsgálatával a megújult tudománytörténeti kutatások a különböző folklorisztikai korszakok metadiszkurzív gyakorlatainak feltárását végezték el.<sup>30</sup> A magyarországi folklorisztika esetében is az utóbbi évtizedek folklorarchívumi kutatásai – egy-két főként

---

26 Ehhez vö. a *NIF (Nordic Institute of Folklore) Newsletter* füzeteit. A NIF 1974-től az 1990-es évekig szervezett a folklorisztikai archiválásról szóló konferenciákat (vö. Honko 2001). Ugyanakkor máshol a kritikák és a gyűjtési módszertan mentén újrastrukturált diszciplínák nem új archiválási technikákhoz, hanem a gondolatától való teljes elhatárolódáshoz vezettek. Az antropológusok, folkloristák és vallásantropológusok különböző archiválási attitűdjéről lásd Mahlamäki (2001: 2–3).

27 Lauri Honko kifejezése (*dead artifacts*), vö. ehhez Anttonen (2013: 159–161) és Gunnell et al. (2013: 173).

28 Az eljárást CollCardnak (*collection card*) nevezték és az 1980-as évek végén kezdték alkalmazni a turkui egyetemen. A nagyszabású indiai Siri eposz rögzítésekor már ezt alkalmazták, melyből később háromkötetes publikáció született (Honko 1998, 2001; Mahlamäki 2001). A CollCard technikával a terepmunka során az adatok rögzítésekor kitértek például az adatközlő émikus klasszifikációjára, a kontextus árnyalt rögzítésére (interjú: hallgatóság nélkül/hallgatósággal/csoportos/autentikus performansz, indukált kontextus, rejtett dokumentáció, aktív közönség, passzív közönség stb.) (Rajamäki 1989). A dokumentálás új eszményeként a sűrű korpuszok (*thick corpuses*) készítését szorgalmazták (Honko 2000b: 21–22); máshol az általuk alkalmazott aprólékos dokumentálást szövegtanográfának (*textual ethnography*) is nevezték.

29 Többen ezért azt javasolták, hogy kutatóink inkább készítsen egy egyszerű, olvasható szöveget, amihez azonban csatolja az előadás/esemény sűrű leírását (Honko 2000a: 36).

30 Például Charles L. Briggs tanulmánya a Grimm testvérek textualizációs eljárásairól (Briggs 1993).

összehasonlító kutatástól eltekintve – gyakorlatilag a 19. századi folklorisztika tudománytörténeti és társadalomtörténeti kontextusainak megértésére fókuszáltak.<sup>31</sup> A folklórarchívumi anyagtól való elfordulás még a történeti folklorisztikát művelőkre is jellemző lett. Annak ellenére, hogy korábbi korszakok folklórjának új típusú vizsgálatához (a történeti antropológiával és a mikrotörténeti kutatásokkal karöltve) bizonyos országokban a folklorisztika gyümölcsözően használta fel a legkülönbözőbb levéltári és archívumi anyagokat (boszorkányperek, egyházi vizitációs jegyzőkönyvek, szentszéki iratok stb.),<sup>32</sup> a történeti anyagok iránt fogékony folkloristák szintén mellőzték a szigorú értelemben vett folklórarchívumokat.<sup>33</sup>

Ahhoz, hogy a 19. század végi, 20. század eleji folklórgyűjtemények a maguk teljes valójában újra kutatásra érdemesek legyenek, a digitális technológiára volt szükség.<sup>34</sup> Az adatbázis mint a számítógépes kultúra sajátos kifejezőeszköze (Manovich 2009 [2001]) új impulzust adott a sok helyen mellőzött és ódivatúnak vélt folklórarchívumi kutatásoknak és folklorisztikai textológiának, s ezáltal az archívumok ismét a kutatások középpontjába kerülhettek. A folyamat jól érzékelhető a vezető folklorisztikai társaságok törekvéseit nézve, ahol egyre markánsabban jelenik meg a folklórarchívumok, gyűjtemények örökségének problematikája.<sup>35</sup> A néprajztudomány mindkét rangos nemzetközi szervezetén belül (ISFNR: International Society for Folk Narrative Research; SIEF: Société Internationale d’Ethnologie et de Folklore) külön munkacsoportok és bizottságok foglalkoznak a kérdéskörrel,<sup>36</sup> a finn Folklore Fellows által szervezett 2009-es nyári egyetem témája a terepmunka és archiválás kapcsolata,<sup>37</sup> a 2015 júniusában megrendezetté pedig a digitális folklorisztika volt, mely utóbbiban két külön szekciót is szerveztek a digitális folklór-adatbázisok és digitális folklór-szövegkiadások kérdésének megvitatására.<sup>38</sup>

A konferenciákon és kongresszusokon kívül számos folyóiratszám és tanulmánygyűjtemény látott napvilágot a témában,<sup>39</sup> a 2010-es évek elején Timothy R. Tangherlini több tanul-

31 Főként 19. századi vagy századfordulós kéziratban maradt gyűjtések és azokhoz kapcsolódó járulékos anyagok (például levelezések) kiadása és feldolgozása Domokos Mariann, Gulyás Judit, Olosz Katalin, Szakál Anna, vagy a 17–18. és 19. század eleji közköltészetet feltáró Küllös Imola és Csörsz Rumen István munkássága által (Bárth 2012: 13–15).

32 A magyarországi történeti folklorisztikai kutatásokról röviden lásd Bárth (2012: 15–17).

33 Persze kivételek mindig akadnak, például Lauri Honko korai munkássága vagy Anna-Lena Siikala kutatásai stb.; lásd Harvilahti (2012: 405).

34 Bár folklórszövegek elemzésére számítógépes eljárásokat az 1960-es évektől alkalmaztak (lásd Voigt 1981), és az archívumok katalógusainak digitalizációjával számos országban az 1970-es évektől foglalkoznak – vö. *NIF Newsletter* 1982 10(4); a számítógép okozta technológiai fejlődés paradigmaváltó szerepe és jelentősége csak az elmúlt 10–15 évben vált egyértelművé a világháló megjelenésével és az online adatbázisok tömeges elterjedésével.

35 ISFNR Vilnius 2013, SIEF Zagreb 2015, a SIEF Working Group on Archives és a lett folklórarchívum, valamint a Network of Nordic and Baltic Tradition Archives szervezésében 2016-ban Rigában tartott „Towards Digital Folkloristics” című konferencia, SIEF Göttingen 2017.

36 Az ISFNR 2005-ben Tartuban megrendezett kongresszusán merült fel először, hogy a népmesék digitális-adatbázis-alapú feldolgozásának, illetve a mesei szöveggyűjtemény világhálón történő terjedésének vizsgálatára egy új albizottságot hozzanak létre, amely végül *Committee for »Folktales and the Internet«* névvel 2009-ben, az Athénban megrendezett kongresszuson állt fel. Lásd erről bővebben: <http://www.isfnr.org/index2.html>. 2013-ban a szintén tartui rendezésű SIEF kongresszuson megalakult a SIEF Working Group on Archives munkacsoport is, ahol kiemelten foglalkoznak a folklórarchívumok digitalizációjával. Lásd <http://www.siefhome.org/wg/arch/index.shtml> (letöltve: 2017. június 6.).

37 Lásd <http://www.folklorefellows.fi/principles-of-fieldwork-and-archiving/> (letöltve: 2017. június 6.).

38 Lásd [http://www.folklorefellows.fi/?page\\_id=2648](http://www.folklorefellows.fi/?page_id=2648) (letöltve: 2017. június 20.).

39 Vö. ehhez például az *Oral Tradition* című folyóirat *Archives, Databases and Special Collections* (2013) különszámát, <http://journal.oraltradition.org/> (letöltve: 2017. január 10.); valamint Holger, Schmitt, Janssen és Schering munkáját (2014).



mányában körvonalazta az általa *számítógépes folklorisztikaként* (*computational folkloristics*) megnevezett új kutatási irányzat feladatait és kihívásait (Abello, Broadwell és Tangherlini 2012; Tangherlini 2014, 2016a). Az utóbbi egy-két évben pedig a korábbi, a digitalizálási projekteket pusztán leíró, azok technikai részleteiről informáló, főként bemutató jellegű tanulmányok helyett egyre többet olvashatunk folklorisztikai adatbázisokon végzett számítógépes elemzésekről is (Abello, Broadwell és Tangherlini 2012; Tangherlini 2013a, 2014, 2016a, 2016b; Kenna, MacCarron és MacCarron 2017).

## Digitális folkloradatbázisok – nemzetközi megközelítési irányok

Írásom célja e fent nevezett új irányzat jegyében az elmúlt 15–20 évben létrejött nemzetközi törekvések összefoglaló bemutatása, kimondottan arra fókuszálva, hogy a digitális folkloradatbázisok a 20. század eleji folklorarchívumi anyaggal és a hozzá kapcsolódó elméleti és módszertani problémákkal hogyan küzdenek meg az új mediális térben, azaz hogyan teszik közzé az anyagokat a digitális szférában, milyen elképzelések és módszerek mentén kívánják megszólaltatni a korábbi gyűjtések *halott leleteit*. Az összefoglalás nem teljes körű, inkább kísérlet arra, hogy néhány fontosabb projekt kapcsán a jellemző irányokat megragadják és rávilágítsanak a digitalizálásban rejlő lehetőségekre és buktatókra.

A folklorisztikai adatbázisokat szemügyre véve rögvest két, egymástól lényegesen eltérő megoldás körvonalazódik, melyek tudománytörténeti előzményekre és az archívumok és gyűjtemények eredeti struktúráira vezethetők vissza. Az egyik a folklorműfajok szerinti digitális adatbázisok gyakorlata, a másik az egy gyűjtő vagy gyűjtési hálózat teljes anyagát digitalizáló projektek.

### *Középpontban a műfaj*

A műfajadatbázisokkal a legelső folklorisztikai textológiai eljárás folytatódik. A *népmesék, balladák, népdalok, népmondák* publikálása a 19. századi műköltészet műfajhierarchiájának mintájára létrehozott szövegkiadásokkal vette kezdetét, majd a 20. század első felében a földrajztörténeti irányzatok katalógusaival tetőzött, hiszen az archívumi folkloranyag rendszerének egyik legfontosabb eleme a műfaj volt.<sup>40</sup> A folklorműfajokat és a műfajiságot ért többirányú kritikák ellenére mindez máig jellemző eljárása maradt a folklorisztikai textológiának.<sup>41</sup> Az egy-egy folklorműfajt középpontba állító adatbázis népszerűségét mutatja, hogy az interneten böngészve nagyobb erőfeszítés nélkül bukkanhatunk rá például a finn és észtrunó,<sup>42</sup> az észtr mondóka, hiedelemszöveg, találós kérdés,<sup>43</sup> az izlandi hiedelemszöveg,<sup>44</sup>

40 A természettudományok mintájára kialakított hierarchikus osztályok gondolatához vö. Tangherlini (2013b: 39–40).

41 A folklorműfajok problematikájához lásd az 1980-as évekig Ben-Amos kritikus összefoglalóját (Ben-Amos 1981), később pedig Finnegan írását (Finnegan 1992: 127–147), továbbá a műfajiság émikus-étikus kategóriájának kérdéséről Ben-Amos (1969), a műfajok instabilitásáról Shuman és Hasan-Rokem (2012: 61–62) írásait. A folklorműfajok recens elméleteihez összefoglalóan legutóbb Shuman és Hasan-Rokem (2012) munkáját.

42 <http://skvr.fi/> és <http://www.folklore.ee/regilaul/andmebaas/> (letöltve: 2017. június 6.), bővebben az adatbázisokról: Saarinen (2001) és Harvilahti (2013).

43 Az alábbi linken érhető el az összes észtr folkloradatbázis: <http://en.folklore.ee/dbases/> (letöltve: 2017. január 10.), továbbá vö. Järvi (2013: 295–296).

44 <http://sagnagrunnur.com/en/> (letöltve: 2017. június 6.), az adatbázis bemutatását lásd Gunnell (2010).

a pánhiszán *ballada*,<sup>45</sup> a szefárd zsidó *népköltészet*,<sup>46</sup> az izraeli *proverbium*,<sup>47</sup> az angol *ponyvaballada*<sup>48</sup> vagy a román *szerelmi ráolvasás*<sup>49</sup> adatbázisokra. E műfajspecifikus megközelítésben egyértelműen a *mese* vezet, létezik holland, flamand, portugál, katalán, örmény, dán, izlandi, német és francia online meseadatbázis is.<sup>50</sup>

A folklórműfajok alapulvételével e műfajadatbázisok a földrajztörténeti irányzatok, s ezzel együtt a komparatív megközelítések célkitűzéseit is megörökítették. A rajtuk végzett elemzések egyik elsődleges célja ugyanis, hogy az összehasonlító szövegfolklorisztika számos kérdését újragondolva a digitális technológia segítségével azokra innovatív módon válaszoljon.<sup>51</sup> Mi az, amire eddig nem volt megoldás, de most, a digitális médium által lehetővé válik? Vitathatatlan előnye az adatbázisoknak, hogy sokkal nagyobb számú szöveg közreadására alkalmasak, mint korábban bármilyen nyomtatott szöveggyűjtemény-sorozat. Ugyan sokan megkérdőjelezik e mennyiségi tényező jelentőségét (mit érünk el azzal, ha 1000 mese helyett 1 millió mesét elemzünk?), a nagy léptékű adatelemzések vagy a számítógépes irodalomtudomány szerint azonban a nagy mennyiségű digitalizált tartalom a kánon, azaz a vizsgált szövegtörzshoz jelentős átalakulást eredményezi majd (Moretti 2000; Jockers 2013).<sup>52</sup> A folklorisztikában mindez azzal is jár, hogy az adatbázisok révén a kevésbé reprezentatív, csonka szövegek, a töredékek és a variánsok is bekerül(het)nek a kutatás körforgásába. Így nem csak a korábbi nemzedékek által előválogatott, a saját kornak és kutatási célnak megfelelő textualizációs eljárás szerint kiadott szövegeket lehet használni, hanem az eddigiekhez képest egy teljesebb s talán kevésbé terhelt korpuszt is (Jockers 2013; Tangherlini és Leonard 2013).

A digitális adatbázis másik nagy előnye, hogy a szövegek többféle rendezését teszi egyszerűen lehetővé, melyben ötvözni lehet korábbi kiadási gyakorlatokat. Nem kell ugyanis eldönteni, hogy régió, település, etnikum, típus, gyűjtő vagy adatközlő szerint közöljük az anyagot, az adatbázisban könnyedén és gyorsan fókuszot válthatunk a folklorisztikai anyag azon alapadatai mentén, amelyek rögzítésre kerültek az adatbázisban. Például az izlandi hiedelemszöveg, a holland mese vagy a finn és észtrunó adatbázisok mindegyikében lehet keresni gyűjtő, adatközlő és helyszín szerint is. A leginkább felszabadító újdonsága az adatbázisoknak a folklorisztika számára azonban az, hogy végre nem kell a szövegeket csupán egyetlenegy kategóriába besorolni. Azokat egyszerűen többféle kategóriához is hozzá

45 <http://depts.washington.edu/hisprom/> (letöltve: 2017. január 10.).

46 <http://sephardifolklit.illinois.edu/> (letöltve: 2017. április 21.), az adatbázis bemutatását lásd Rosenstock és Bistué (2013).

47 *Az Israeli Proverb Index Project* (IPIP) jelenleg online formában nem elérhető, az adatbázis bemutatását lásd Belinko és Kats (2014).

48 <https://ebba.english.ucsb.edu/> (letöltve: 2017. január 18.), az adatbázis bemutatását lásd: Fumerton és Nebeker (2013). Illetve az oxfordi Bodleian Library vállalkozásában a *Broadside Ballads Online* elnevezésű projekt, melyben a kora újkortól a 20. századig gyűjtöttek össze angol ponyvaballadákat; <http://ballads.bodleian.ox.ac.uk/> (letöltve: 2017. január 18.).

49 <http://cds.library.brown.edu/projects/romanianCharms/> (letöltve: 2017. január 18.), az adatbázis bemutatását lásd Golopenția (1997).

50 Theo Meder összefoglalása az általa ismert online elérhető meseadatbázisokról (Meder 2014b: 2), <http://www.isfnr.org/files/CommitteeInternet.pdf> (letöltve: 2017. június 6.).

51 Az összehasonlító módszer és a földrajztörténeti irányzat(ok) digitális adatbázisoktól független folklorisztikai revitalizációjához lásd Dégh Linda bevezető tanulmányát (Dégh 1986) és a *Journal of Folklore Research* „The Comparative Method in Folklore” című különszámának tanulmányait (1986), valamint Virtanen (1993) és Wolf-Knuts (2000) írásait. Legutóbb Frog mutatta be tanulmányában, hogy a földrajztörténeti módszerek miként alkalmazhatók termékenyen a kurrens folklorisztikai kutatásban (Frog 2013: 23–30).

52 A probléma elméleti alapfeltevéseiről lásd bővebben a tanulmány „Elméleti keretek” című alfejezetét.

lehet rendelni, s így a közöttük lévő multidimenzionális kapcsolati háló végre megragadható (Meder 2014a; Abello, Broadwell és Tangherlini 2012; Tangherlini 2014; Schmitt 2014). Sőt megfelelő digitális textualizációt követően nemcsak a teljes szövegben történő keresés biztosított, hanem különféle szövegbányászati eszközökkel az is elérhető, hogy az összehasonlítást segítő multidimenzionális kategóriarendszert ne csupán a kutató/kutatók állítsák elő és határozzák meg, hanem magából a szöveganyagból algoritmusok segítségével a programok mutassák fel azt, hogy milyen tulajdonságok alapján lehetnek egyáltalán hasonlóak bizonyos szövegek.<sup>53</sup>

A digitális technológiával az adatok térképre vetítésének egyik sokat kritizált pontja is megoldódik. Korábban ugyanis az analóg néprajzi atlaszok megoldhatatlan nehézsége volt a diakrón adatok szinkrón térképen való kimerevített ábrázolása, a legtöbb kritika éppen emiatt érte a kartográfiai módszert (Munk és Jensen 2014: 40–41). A digitális térképek az idő és a tér dimenzióját azonban képesek egyszerre ábrázolni. A holland mese- és az izlandi hiedelemszöveg-adatbázisokban is kiválaszthatjuk, hogy épp melyik periódus szövegeit szeretnénk a térképen látni.

Az új médium sokféleképpen segíti és reformálja meg tehát a szövegfolklorisztikai kutatásokat, amiktől számos eredményt remélhetünk. Ugyanakkor a műfaji megközelítésben előállított adatbázis koncepciója több elméleti és módszertani kérdést is felvet. A műfajadatbázisok számára, úgy tűnik, gyakorlatilag előre adottak a korpuszok, legalábbis az archívumok műfaji cédulakatalógusai, illetve a műfajalapú folklórszöveg-kiadások ezt sugallják. Néhány adatbázis valójában csupán egy korábbi, lezárt korpusz digitalizált változata. Ilyen például a finn runoadatbázis, amely a *Suomen Kansan Vanhat Runot* (The Ancient Poems of the Finnish People) 34 kötetes (1908–1948, 1998) könyvsorozatát tette elérhetővé és ezáltal több mint 89 000 szöveget digitálisan kutathatóvá.<sup>54</sup>

Komplikáltabb esetről van szó azonban akkor, ha a készítő az adott műfajt reprezentáló fontosabb kötetekből válogatva állítanak össze egy adatbázist. Jó példa erre az izlandi hiedelemszöveg-adatbázis, a *Sagrgrunnur*. Korábban semmiféle mutatója, gyűjteménye vagy katalógusa nem volt az izlandi hiedelemszövegeknek, a kiadások pedig egyáltalán nem hatkozták a nemzetközi párhuzamokat, típusokat. Tery Gunnell diákjaival 2010-ben egy kb. 10 000 szövegre mutató digitális adatbázist publikált izlandi hiedelemszöveg-gyűjteményekből. Fontos azonban, hogy az előzőekben példaként említett adatbázisokhoz képest itt nem közölnek szövegeket, hanem valójában csak a szövegek izlandi (és néhány esetben angol nyelvű) tartalmi kivonatát kapjuk meg egy angol nyelvű keresőfelülettel, kulcsszavas keresési lehetőséggel, amivel tulajdonképpen egy sohasem létezett izlandi hiedelemszöveg-katalógus innovatív pótlását végezték el, teljesen kikerülve ezáltal a különféle kiadványokból is fakadó textualizációs problémákat (Gunnell 2010).

Bár a korábbi könyvkiadások és katalógusok segítségével viszonylag problémamentesen ki lehet választani azokat a szövegeket, amelyek egy adott műfajadatbázis gyűjtőkörébe tartoz(hat)nak, sok esetben merülhetnek fel dilemmák a korpuszépítés körül, hiszen a

53 A szövegbányászati módszerekről lásd bővebben a tanulmány „Szövegbányászat (*text-mining*) és hálózati elméleti módszerek (*network theory methods*)” című fejezetét.

54 Az adatbázis *Eesti regilaulude andmebaas* (The Database of Estonian Oral Poetry) elnevezésű észt párjában – amely tulajdonképpen a finn mintájára jött létre és jelenleg kb. 80 000 szöveget tartalmaz – azonban már kéziratos és kiadott szövegeket is publikálnak, sőt amennyiben egy szöveg két filológiai változatban ismert (kézirat és publikáció), válhatunk a két átírás között.

különböző kultúrák és korszakok (és folklorisztikai gyakorlatok) eltérő műfajrepertoárt működtetnek, az adatbázisok pedig általában egy a 19. század végén, 20. század elején megkonstruált és kimerevített műfajkonceptiót alkalmazva szelektálnak a szövegek között, annak ellenére, hogy időben tág periódust ölelnek fel. Ki dönti el, hogy egy 17. századi vallásos-mágikus szöveg ima vagy ráolvasás?

Nehezítő körülmény, hogy a műfajadatbázisok annak érdekében, hogy minél több szöveget tartalmazzanak az adott műfajból, általában túllépnek a fizikálisan létező szöveggyűjtemény(ek) vagy archívum(ok) keretein<sup>55</sup> – így az adatbázisoknak meglehetősen heterogén forrás csoportot kell megjeleníteniük és kezelniük. A holland meseadatbázisban például szerepelnek szövegek 16. századi kéziratokból, 20. századi folklorisztikai gyűjteményekből, újságokból, ugyanakkor az internetről is (Meder 2014a).

További kritikaként hozható fel, hogy a mesterséges műfajok alapján elképzelt adatbázisok az orális kultúrát szétdarabolják, ezért nem alkalmasak a szóbeliség komplex megértésére, egy-egy lokalitás szöveganyagának vizsgálatára, sőt valójában az újragondolt összehasonlító szövegfolklorisztikai kutatásokat is gátolják, hiszen a külön-külön működő adatbázisok nem engedik láttatni a motívumok és témák műfajokon átívelő sokszínű viszonyrendszerét. Jó példa erre az észt folkloristák esete, akik az internet nyújtotta lehetőségeket a kezdetektől jól kiaknázva mára már több mint húszféle digitális műfajadatbázissal büszkélkedhetnek. Jelen pillanatban azonban az egyik legnagyobb kihívás számukra az, hogy e műfajadatbázisokat hogyan integrálják egy közös digitális repozitóriumba, biztosítva ezzel a különféle virtuális archívumokban történő együttes keresést és kutatást (Köiva 2003; Järv és Sarv 2014: 55–56; Järv 2013: 295–296).

Felismerve a problémát, jelenleg egyre több, inkább átfogó műfaji kategóriákban gondolkozó, tematikus folkloradatbázis készül. A holland meseadatbázis például lassan már csak az elnevezésében őrzi „mese” jellegét. Theo Meder és munkatársai kiterjesztett célja ugyanis a teljes holland népi próza dokumentálása, ezért ma már az anekdotától és mondától kezdve a hiedelemszövegen át a személyes történetig sokféle anyagot rögzítenek a *Nederlandse VolksverhalenBank*-ban (Meder 2010, 2014a). Az oslói egyetem norvég mágia-adatbázisát háromféle forrástípus feldolgozásából állították össze: kora újkori grimoárokat (mágiatan-könyveket) és boszorkánypereket, valamint 19. századi és 20. század elején gyűjtött hiedelemszövegeket tesznek közzé egy közös felületen,<sup>56</sup> a portugál legenda-adatbázis pedig egyszerre tartalmaz szakrális, történeti és városi legendákat, eredetmondákat és a természetfeletről szóló történeteket.<sup>57</sup>

Ilyen fokú heterogenitásnál kérdéses, hogy az adatbázisok e sokféle forrás kontextusát hogyan tudják rögzíteni és megjeleníteni, hogyan tudják a különböző típusú szövegek által felvetett textualizációs problémát megoldani.<sup>58</sup>

55 A korpuszépítés dilemmái a digitális bölcsészettudomány egészére jellemzőek, magyar irodalomtudományi példa Labádi Gergely tanulmánya a 19. századi „magyar regény” korpuszának összeállítására vonatkozóan (Labádi 2014).

56 *Trolldomsarkivet*, Norsk Folkeminnesamling; <http://www.hf.uio.no/ikos/english/services/norwegian-folklore/magic-in-norway/>. Keresni és használni az adatbázist csak norvég nyelven lehet: <http://www.edd.uio.no/ikos/trolldom.html> (letöltve: 2017. szeptember 28.).

57 <http://www.lendarium.org/> (letöltve: 2017. június 20.).

58 A problémához vö. a tanulmány „Technika. Platformfüggetlenség, átjárhatóság, fenntarthatóság” című alfejezetét.

A folklóradatbázisok másik fő megközelítési iránya, amikor egy prominens gyűjtő teljes gyűjteményének, hagyatékának digitalizálása a cél, ami – az európai archívumok esetében – sokszor egybeesik az intézmény fontos, alapító vagy meghatározó törzsállományával,<sup>59</sup> ennél fogva prioritást élvez régisége miatt is.<sup>60</sup> Ez a megközelítési mód azonban nemcsak a fizikai objektumok digitális megmentése miatt lehet indokolt, hanem már csak abból kiindulva is, amit a tanulmány elején jeleztem, hogy tulajdonképpen kutatási kérdésekhez kapcsolható járulékos anyagok alkotják a folklóarchívumok nagy részét, s ezekben az esetekben e komplex anyag digitalizálása a cél. Ezek közül két fontosabb példát szeretnék kiemelni: az Evald Tang Kristensen (1843–1929) dán folklorista hagyatékát feldolgozó *Danish Folklore Nexus* és *ETKSpace* címet viselő vállalkozások elsősorban Timothy R. Tangherlini nevéhez köthetők, a *WossidDiA*, azaz a *Digital Wossidlo Archive* pedig Christoph Schmitt vezetésével Richard Wossidlo (1859–1939) mecklenburgi folklorista örökségét tette digitálisan elérhetővé.<sup>61</sup> A nagyjából ugyanabban a periódusban létrejött két gyűjtemény közötti legszembevetőbb hasonlóság azok szinte beláthatatlan mérete. Kristensen közel ötven év leforgása alatt kb. 4000 adatközlőtől több mint negyedmillió szöveget (vegyesen: balladákat, népdalokat, népmeséket, hiedelemszövegeket, proverbiumokat, ráolvasásokat, mondókákat/rigmusokat, népi játékokat, vicceket, népi gyógyászathoz kapcsolódó, és a mindennapi élet egyéb területeit érintő leírásokat) rögzített. Összességében 12 000 helynév, 60 000 gyűjtési adat, több mint 24 000 oldalnyi kézirat alkotja a gyűjteményt (Abello, Broadwell és Tangherlini 2012: 63–65; Tangherlini 2013b). Wossidlo etnográfiai és dialektológiai gyűjtései során szintén rengeteg témára (például népszokások és hiedelmek, tenger menti vidéki munkák és mindennapi élet, etnobotanika, szexualitás stb.) és műfajra (mese, vicc, anekdota, mondóka, proverbium) figyelt, a rostocki egyetem néprajzi intézetében található archívum Wossidlo részlege nagyjából 2 millió adatból és több ezer kéziratból tevődik össze (Holger, Schering és Schmitt 2014: 64–65).

A két gyűjtemény természetete és a digitalizálás célkitűzései azonban más-más megvalósítást eredményeztek. Kristensen a 19. századból ismert folkloristák közül hihetetlen mennyiségű gyűjtésén túl nagyfokú precizitásával emelkedik ki (variánsok, adatközlők életkörülményei, élettörténetei, mindennapi élet dokumentálása stb.). Hagyatéka a sokféle forrástípus tekintetében is páratlan, hiszen egyszerűen maradtak fent saját kéziratai, pizkozatai, több ezer oldalas terepnaplója, memoárja, a hozzá íródott levelek csatolmányaiakkal, azaz a beküldött kéziratok lokális folklórgyűjtemények sokaságával. Mindehhez hozzáátéve az általa sajtó alá rendezett és kiadott műveket, valamint tematikus mutatókat, kibontható a 19. század végi, 20. századi eleji gyűjtési és kiadási módszertan, a szövegek textualizációjának gyakorlata. A *Danish Folklore Nexus* csak egy, mondhatni kis korpuszba enged bepillantást, öt adatközlő kb. 500 szövegét dolgozza fel, viszont azt teljes komplexitásában. Legnagyobb érde-

59 Ugyanakkor online adatbázisok szétszóródott hagyatékokat is összekapcsolhatnak a virtuális térben, lásd például Adolf Spamer vagy az Opie házaspár hagyatékát feldolgozó projekteket. (Seifert és Keller 2014; Bishop 2013). <http://www.opieproject.group.shef.ac.uk/> (letöltve: 2017. január 18.), <http://www.opieproject.group.shef.ac.uk/about-collection.html> (letöltve: 2017. január 18.).

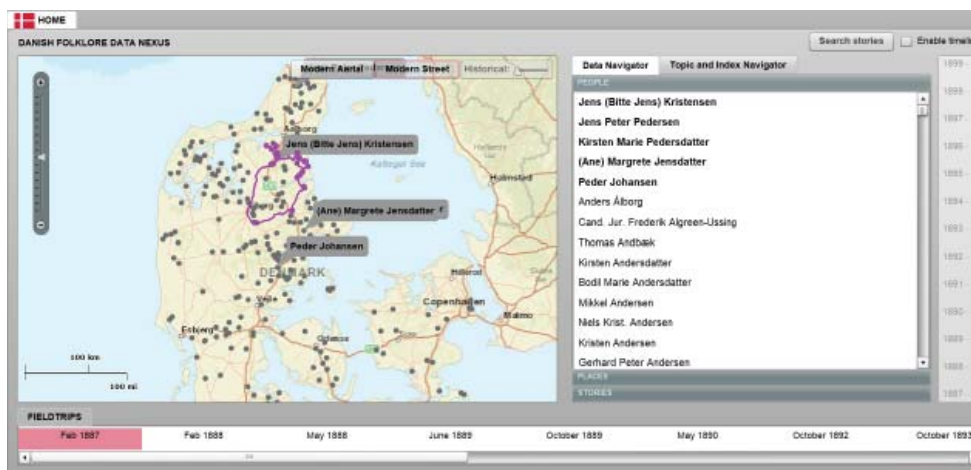
60 Vö. pl. a magyar Kálmány Lajos-hagyaték digitalizálását (Granasztói 2008).

61 <http://etk-space.scandinavian.ucla.edu/macroscope.html>; <https://apps.wossidia.de:8443/webapp/run> (letöltve: 2017. június 27.).

me, hogy a gyűjtőutat, a gyűjtést és az ezekből létrehozott különböző szövegeket szimultán szemléltethetjük. Praktikus ablakrendszerben egymás mellé helyezhető a kéziratos szkennelt képe, annak átirat változata és angol fordítása, de ezek mellé akár a Kristensen által sajtó alá rendezett verziót is megjeleníthetjük. Mindezt a gyűjtési terepmunka kontextusában is elhelyezték, ugyanis digitális térképre vetítve nyomon követhetjük Kristensen egyes gyűjtőútjait és állomásait, amihez ráadásul korabeli, 19. századi térképeket használtak. A digitális felület ekképpen lehetővé teszi azt, hogy a kutató ugyanazt az anyagot különböző perspektívából szemlélje (pl. mikro és makro szint), vizsgálata közben pedig bármikor könnyedén váltszon azok között (Tangherlini és Broadwell 2014).

Mivel Kristensen hagyatéka nemcsak komplexitása miatt fontos, hanem óriási anyagmennyisége miatt is, a másik felületet, az *ETKSpace*-t ennek kezelésére hozták létre. Itt a teljes, kiadott anyag digitalizált változatában biztosított a keresés és a szövegelérés, ugyanakkor kevesebb metaadat közül választhatunk. A több mint 30 000 szöveget tartalmazó korpuszban Kristensen saját mutatói, valamint egy ezt kiegészítő, Tangherlini korábbi kutatásainak eredményét összegző kategóriarendszer, kulcsszavak és helyszínek alapján kereshetünk.

1. ábra. A *Danish Folklore Nexus* online felülete, Evald Tang Kristensen 1887-es gyűjtőútja térképen ábrázolva

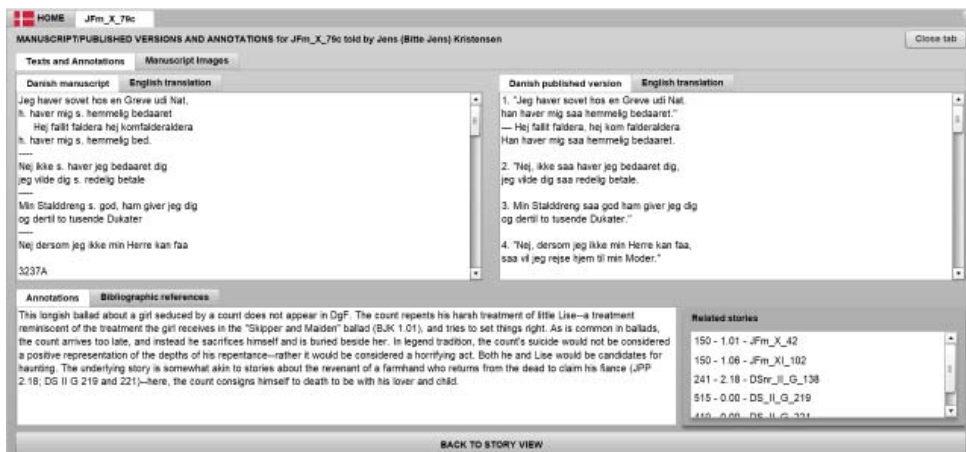


Forrás: [etkspace.scandinavian.ucla.edu](http://etkspace.scandinavian.ucla.edu)<sup>62</sup>

62 <http://etkspace.scandinavian.ucla.edu/danishfolklore/#> (letöltve: 2017. szeptember 28.).

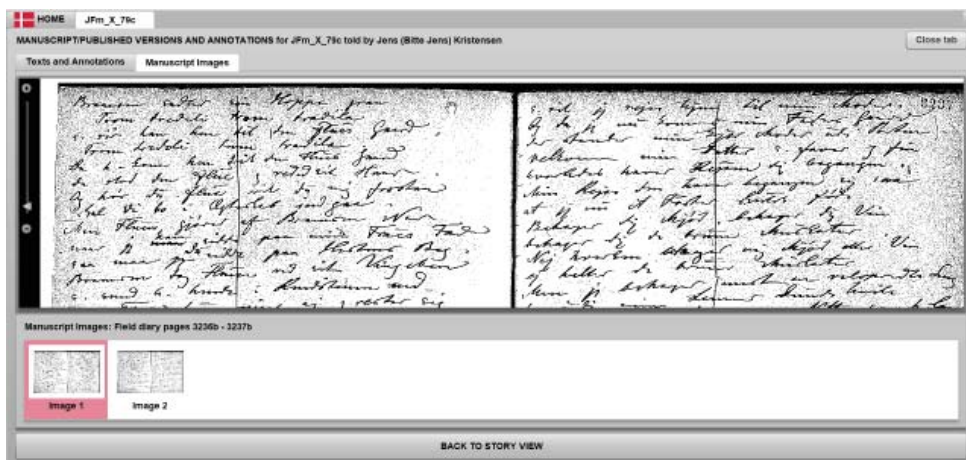


2. ábra. A Danish Folklore Nexus online felülete, a különféle szövegváltozatok párhuzamos olvasását biztosító ablakrendszer



Forrás: etkospace.scandinavian.ucla.edu<sup>63</sup>

3. ábra. A Danish Folklore Nexus online felülete, az eredeti dokumentum szkennelt képe



Forrás: etkospace.scandinavian.ucla.edu<sup>64</sup>

A Tangherlini-féle koncepció tehát egyrészt a gyűjtés kontextusára, a gyűjtésben részt vevő emberek és szövegek közötti különböző viszonyok és a korszak folklorisztikai-textológiai gyakorlatának feltárására alkalmas felületet hozott létre, másrészt Kristensen nyomtatásban is megjelent hatalmas mennyiségű szöveganyagának automatizált digitalizációjával (OCR) a teljes korpuszt tette kereshetően elérhetővé. Az alkotók által megvalósított felületek ekképpen jól illeszkednek a tanulmány elején említett megújult folklorisztikai tudománytörténeti

63 <http://etkospace.scandinavian.ucla.edu/danishfolklore/#> (letöltve: 2017. szeptember 28.).

64 <http://etkospace.scandinavian.ucla.edu/danishfolklore/#> (letöltve: 2017. szeptember 28.).

irányba, ugyanakkor a szövegek mennyisége és koherenciája – sok műfajadatbázissal ellentétben – egyaránt alkalmas összehasonlító szövegfolklorisztikai kutatásokra, s ebből következően a 19. századi vidéki Dánia történeteinek mélyebb megértésére is. A *Danish Nexus* és az *ETKSpace* ellen az eddig említett pozitívumok mellett kritikaként hozható fel, hogy a gyűjtemény forrásadottságait a digitális felület nem láttatja kellőképpen. A honlap nem nyújt semmiféle bevezetést, tájékoztatást arról, hogy tulajdonképpen milyen forrásokat dolgoz fel, a korpusz mekkora részét és hogyan tették elérhetővé az alkotók. Az oldalra fellépő felhasználók Tangherlini nyomtatásban megjelent könyvéből tudják a honlapon is használt rövidítéseket feloldani és némiképp értelmezni (Tangherlini 2013b: xviii–xx).

Az előzőhöz képest a *WossiDiA* Richard Wossidlo anyagát az archívum kontextusában, az archívum rendszerének komplexitásában akarja megmutatni. A *WossiDiA* legnagyobb részét Richard Wossidlo kézzel írt cédulái alkotják, amelyeken főként saját vagy gyűjtőitől kapott mecklenburgi gyűjtéseit adatolta. Wossidlo nemcsak gyűjtési adatokat cédulázott ki, hanem különböző publikált forrásból (folyóiratok, könyvek, újságok) is gyűjtötte a mecklenburgi adatokat és azok német néprajzi, kisebb mértékben szláv és skandináv párhuzamait. Az archívum második legfontosabb egysége a gyűjtői levelezéskorpusz. A gyűjtemény igazi jelentőségét azonban maga az archívumi rendszer adja, mely a különféle keresztthivatkozásokkal egy gazdag és sokrétű kulturális rendszer szemiotikájának dokumentálását kísérte meg. A digitalizáció első fázisaként a teljes hagyatékot beszkennelték (az összes cédulát és az összes levelet), majd olyan összetett irányított hipergráf modellt és keresési felületet hoztak létre hozzá, mely a különböző dokumentumok közötti lehetséges kapcsolódási pontokat egyszerre képes tárolni és megmutatni. Az alkotók célja az volt, hogy egy „intelligens digitális archívumban rekontextualizálják az elveszett kapcsolatot a kutatók, illetőleg a gyűjtők és hálózatuk között” (Holger, Schering és Schmitt 2014: 84).

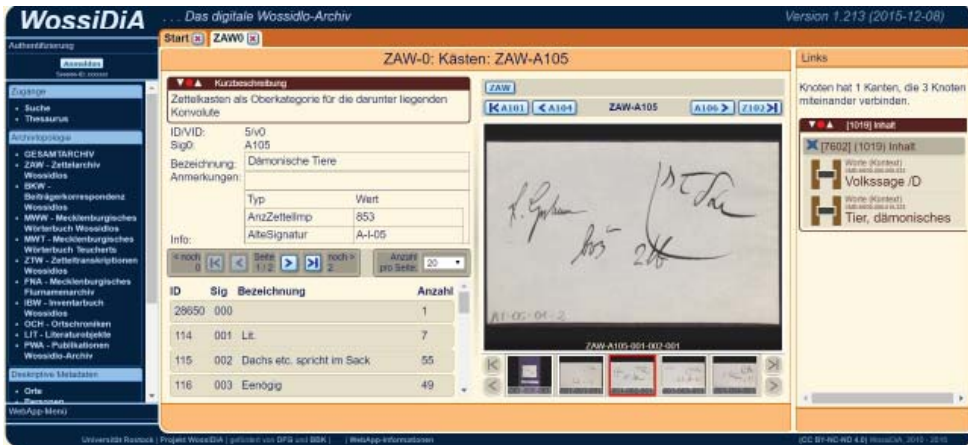
4. ábra. A *WossiDiA* online felülete az eredeti cédulakatalógussal



Forrás: apps.wossidia.de<sup>65</sup>

65 <https://apps.wossidia.de:8443/webapp/run> (letöltve: 2017. szeptember 28.).

5. ábra. A WossiDiA online felülete, egy digitalizált cédula megjelenítése az adatbázisban



Forrás: apps.wossidia.de<sup>66</sup>

Az adatbázis létrehozói szerint ugyan a *WossiDiA*-t kifejezetten a Wossidlo-archívumra tervezték, ugyanakkor az általuk kidolgozott modell általában is alkalmas lehet komplex etnográfiai gyűjtemények digitális feldolgozására és megjelenítésére. Hozzá kell azonban tenni, hogy a *WossiDiA* esetében azért tudták mindezt megvalósítani, mert már készen kapták Wossidlo gazdag kategóriarendszerét és kereszthivatkozásait. Egy elméleti és gyakorlati szempontból is befejezett és lezárt archívumi rendszert emeltek át a digitális szférába, az új technológia pedig azt teszi lehetővé, hogy a rendszer elemei között további kapcsolódási pontok is felszínre kerüljenek. Ugyanakkor a digitális közreadás ebben a megközelítésben nem „írja felül” a már létező kategóriarendszert, az indexeket és a mutatókat, így a Wossidlo által lefektetett alap egyben végérvényesen meghatározza az adatbázist is, annak esetleges hibáival együtt. A legtöbb folklórarchívum ráadásul nem ilyen szerencsés. Ahogy korábban utaltam rá, analóg mivoltukban az archívumok sohasem értek az anyag rendezésének a végére, a *WossiDiA*-féle megoldás előfeltétele viszont egy jól strukturált, rendezett folklóranyag.

A digitális archívumok és elektronikus szövegkiadások egyik legfőbb célja, hogy az adatokat és szövegeket azok *kontextusaiba visszahelyezve* közölje (McGann 2010). Mit jelent azonban mindez a folklórszövegek nézőpontjából? A folklórszövegeket nyilvánvalóan soha nem tudjuk visszahelyezni azok eredeti, elsődleges elhangzási kontextusába, azonban e szövegeknek még számtalan egyéb kontextusa is van (Anttonen 2013), az izgalmas és komplex feladat pedig elsődlegesen éppen ez, hogy a sokféle kontextus közül melyikbe vagy melyekbe próbálják az adatbázis megálmodói visszahelyezni az adatokat. A műfajadatbázisok a tér és idő, valamint sok millió hasonló szöveg; a *Danish Nexus* a gyűjtés, a gyűjtő, az adatközlő és a 19. századi Dánia történelmi és földrajzi viszonyrendszerének; a *WossiDiA* pedig az archívumi struktúra, a gyűjtemény kontextusába helyezte a szövegeket. Digitális létmódjukkal ugyanakkor egy újabb kontextusba, a digitális adatbázisok világába is bekerülnek, ahol az átjárhatóság és hosszú távú fenntarthatóságon van a hangsúly.

66 <https://apps.wossidia.de:8443/webapp/run> (letöltve: 2017. szeptember 28.).

A fentiekből kiindulva tulajdonképpen egy harmadik vonatkozó kérdésről is szót kell ejteni, mégpedig arról, hogy a műfaj- és gyűjteményközpontú adatbázisok hogyan integrálhatók teljes folklórarchívumok digitalizálási elképzeléseibe.<sup>67</sup> Hogyan és hol érnek össze, egyáltalán hogyan lehet őket összekötni? Tanulmányom nem tud részletesen kitérni a folklórarchívumok általános digitalizációjának kérdésére, hiszen fókuszában a folklóradatbázisok állnak, néhány dolgot azonban szükséges érinteni, legfőképpen azért, mert jellemző módon a folklorisztikai adatbázisok egy-egy folklórarchívum részei, és fentarthatóságukat is ezen intézmények tudják biztosítani.

Az elsősorban 19. századi angol irodalommal foglalkozó Jerome McGann szerint az archiválás egyik nagy tudománytörténeti tanulsága, hogy az emberiség mindig elbukik annak eldöntésében, mi érdemes a megőrzésre. A lassan teljes egészében digitalizálódó kultúra pedig sokakban azt az illúziót kelti, hogy a virtuális archívum teljes mértékben leképezi a fizikai valójában létező archívumot, ahova éppen ezért már felesleges is elmenni (McGann 2014: 42–45). Bár a teljes anyag digitalizálása és közzététele sok szempontból nem is célja a folklórarchívumoknak – etikai és jogi kérdésekből, valamint erőforrás hiányából fakadóan –, McGann gondolatait érdemes megfontolni. Mit szükséges digitalizálni? A legkorábbi? A legteljesebbet? A jellemzőt, az egyedit vagy a foszladozót? Sok esetben megfigyelhető az is, hogy a bizonytalan megmaradású digitális prezerváció miatt az eredeti objektumok állagmegóvása elmarad, holott a digitális kép pusztán numerikus kód, ami ezerszer sérülékenyebb, mint a 19. századi papír vagy az ókori pergamen. A *WossiDiA* kétmilliónyi szkennelt cédulája láttán jogosan merülhet fel kérdésként, hogy mi egy archívum, egy gyűjtemény digitalizációjának elsődleges feladata. Érdemes a sokszor téves, pusztán pársoros hivatkozást tartalmazó cédulákat digitalizálni csupán azért, hogy valaki otthonról is elérje? Vagy koncentráljon egy folklórarchívum arra, hogy anyagának tartalmát minél pontosabban feltárja, digitális segédletekkel és mutatókkal pedig biztosítsa a felhasználók számára a minél részletesebb adatkinyerést (Granasztói 2008: 126–127)? Ugyanakkor mindez sok esetben a folklórarchívumok korábban bemutatott rendezetlensége miatt valójában csak a tartalom teljes digitalizációjával érhető el, ahogy a következőkben is látni fogjuk.

### A digitális tartalom előállítása

A folklorisztikai archiválás rövid tudománytörténeti áttekintése, valamint a digitális folklóradatbázisok tömör összefoglalója után lássuk tehát, hogy a gyakorlatban hogyan valósíthatók meg az elképzelések, a digitális bölcsészettudomány milyen megoldásokat kínál, milyen új problémákat vet fel, és hogyan kiviteleztek az elképzeléseket a folkloristák mind ez idáig.

---

<sup>67</sup> Az ész és az ír folklórarchívumok digitalizálásához lásd Järv és Sarv (2014); Ógáin (2013). Magyar vonatkozásban az Etnológiai Archívum fejlesztéséről Granasztói (2013).

A digitalizációs törekvések egyik legfontosabb mozzanata a számítógépes programok által olvasható (szöveg)adat előállítás. Hogyan teremtünk erre forrást? Hogyan tudjuk megszólaltani az archívumban fekvő többmilliónyi szöveget, amelyek katalogizálása sem fejeződött még be? A kéziratok digitális átírása drága és időigényes, amit általában a néhány szakemberrel működő folklórarchívumok önerőből feltehetően sohasem lesznek képesek véghez vinni. Megoldásként erre egyre divatosabbak a *citizen science* vagy másképpen *crowd science*, *crowdsourcing*, *civic science* módszerek (önkéntes adatszolgáltatáson alapuló tudomány, közösségi együttműködés).<sup>68</sup> A *közösségi tudomány* az adat előállításába vagy akár magába az adatgyűjtés folyamatának bizonyos fázisába önkéntes civileket von be. Az önkéntes munka a társadalmi gyűjtőhálózatok révén már ismerősen cseng a folklorisztikában, s úgy tűnik, e megújult változata is hatékonyan működtethető. A folkloriszövegek rögzítésekor önkéntesekre a legnagyobb szükség ott van, ahol az automatikus optikai karakterfelismerő programok (OCR) nem használhatók (pl. mert kéziratos dokumentumokról van szó), vagy használhatók ugyan, de folyamatos manuális ellenőrzésre szorulnak (pl. mert a tipográfia, illetve a nyelvi sajátosságok miatt folyamatosan hibázik a program).

Az ír National Folklore Collection *The Schools' Collection* projektje jó példa a közösségi tudomány gyümölcsöző alkalmazására. A *The Schools' Collection* több mint 5000 ír általános iskola nagyjából 50 000 tanulójának gyűjtése az 1937 és 1939 közötti időszakból, ami kb. 740 000 oldalnyi kéziratot jelent (Briody 2007: 260–270).<sup>69</sup> A dublini egyetem keretein belül működő folklórarchívum *Meitheal Dúchas.ie: Community Transcription* elnevezésű, 2013-ban indított kezdeményezése során szűk négy év alatt több mint 17 000 oldalnyi ír és 64 000 oldalnyi angol nyelvű folklórgyűjtést írtak át önkéntesek (ami a teljes kéziratos anyag ír részének 18%-a, míg az angol nyelvűnek 19%-a).<sup>70</sup> Hasonló kampányt alkalmazott a lett folklórarchívum is (Latviešu folkloras krātuve) *Simtgades burtnieki* (A centenáriumi varázslói/tudósai)<sup>71</sup> című programjával, amikor csatlakozott az ország százéves fennállásának önkéntes civil mozgalmi előkészítéséhez.<sup>72</sup> Az archívum 2016 júniusában indította el kéziratainak átírási lehetőségét, hat hónap alatt pedig majdnem 30 000 oldalnyi szöveget írtak át

68 <http://www.oszk.hu/civic> (letöltve: 2017. január 9.). A *crowdsourcing* módszerek általános összefoglalását és történeti bemutatását lásd Riesch (2015), a módszerek digitális bölcsészettudományi lehetőségeiről összefoglalóan Terras (2016), a digitális szövegkiadásokban betöltött szerepükről kritikusan Shillingsburg írását (2016).

69 <http://www.duchas.ie/en/info/cbe> (letöltve: 2017. január 18.).

70 Lásd <https://www.duchas.ie/en/meitheal/> (letöltve: 2017. október 4.). A laikus résztvevők motivációit, gyorsaságukat és eredményességüket vizsgáló tudományos összefoglalás is készült már: *Crowdsourcing Motivations in a GLAM Context: A Research Survey of Transcriber Motivations of the Meitheal Dúchas.ie Crowdsourcing Project*, <http://digitalirishheritage.com/dissertation/uncategorized/data-visualizations-of-qualitative-data/> (letöltve: 2017. január 9.). A digitális disszertáció bő bibliográfiát és hivatkozást kínál a crowdsourcing-ot használó portálokról és a jelenséget vizsgáló tudományos írásokról.

71 <http://lv100.garamantas.lv/en> (letöltve: 2017. január 9.).

72 Lettország 2018-as centenáriumi tervezetéhez lásd: <http://www.latvia.eu/latvias-centenary> (letöltve: 2017. január 9.).



civilek. A regisztrált felhasználók a gyűjtemény anyagának megfelelően 11 nyelv (lett, latgál, livóniai, litván, észt, orosz, belorusz, jiddis, roma, lengyel, német) közül választhatnak. Az önkéntesség a lett folklórarchívum digitális felületén ettől függetlenül is működik, itt akár hanganyagról is lejegyezhetnek a közösségi tudomány elkötelezett hívei.<sup>73</sup>

Nyilvánvalóan nem minden esetben alkalmazhatóak a fentiek, hiszen a 19. századi vagy korábbi korszakok kézírásának átírásához bizonyos filológiai előképzettség szükséges. Az európai folklórarchívumok 20. században keletkezett hatalmas korpuszainak nagy részét azonban többnyire iskolás gyerekek, tanulók, tanárok vagy falusi értelmiségiek jól olvasható kézírásos vagy gépírási szövegei képezik.<sup>74</sup>

A szövegátírásokat a folklorisztikai textológiában jártas folkloristáknak vagy filológusoknak ellenőrizniük kell (és ellenőrizik is az ír és a lett esetben is), de még így is hatalmas potenciál rejlik a közösségi tudományban, mely természetesen nem csak szövegátírásra használható (hanem többek között adatgyűjtésre, adatellenőrzésre, fényképfelismerésre, adatok lokalizálására stb.). A lett kampányban például az átíráson túl engedélyezik az önkénteseknek azt is, hogy bármilyen általuk létrehozott kulcsszót adjanak az átírt szöveghez. Emellett a közösségi tudomány nem pusztán ingyenmunkaerőt jelent az adatfeldolgozással küszködő közintézmények számára. A civilek bevonásával a múzeumok, kézírattárak növelik társadalmi szerepüket és hasznosságukat, a tudomány elefántcsonttornyából kilépve interaktív módon teszik a civileket a tudástermelés részesévé. A nemzeti, népi, lokális és különféle kulturális örökségek iránti megnövekedett civil érdeklődés, a korábbi társadalmi gyűjtőhálózati tőke a folklórarchívumok esetében különösen működőképessé teheti a participatorikus civiltudományi kezdeményezéseket, melyek a – sokszor parttalanná vált – hagyományos, önkéntes néprajzi gyűjtőmunka megújulását hozhatják.<sup>75</sup>

### *Technika. Platformfüggetlenség, átjárhatóság, fenntarthatóság*

Az adat előállítása tehát időigényes, fáradtságos folyamat, ráadásul még mindig az egyik legdrágább művelete a digitalizációnak – még ha bizonyos esetekben civiltudományi kezdeményezésekkel lehet is csökkenteni e műveletek költségeit. Ebből kifolyólag a hosszú távú megőrzés és a széles körű felhasználhatóság az egyik elsődleges kívánalom, a digitális bölcsészettudomány formálódó paradigmájában azonban itt érzékelhető a legtöbb bizonytalanság. Vajon mi az a formátum, ami tíz év múlva sem évül el? Mire kell egyáltalán előkészíteni az adatot? Hogyan és hol találják meg a különböző bölcsészettudományi diszciplínák a közös nevezőt a digitális textológiában, a digitális adat-előkészítésben, biztosítva ezáltal az átjárhatóságot?

<sup>73</sup> <http://garamantas.lv/en> (letöltve: 2017. január 9.).

<sup>74</sup> A fentiek példáján felbuzdulva önkéntelenül is eszünkbe juthat, hogy ugyanezt a metódust kiválóan lehetne alkalmazni a budapesti Néprajzi Múzeum Etnológiai Archívumában található több mint 10 000 oldalnyi, 1912 és 1921 között keletkezett Folklore Fellows kéziratos anyagának feldolgozásakor, hiszen a szövegeket iskolások gyöngybetűkkel, a legtöbb esetben jól olvashatóan jegyezték le.

<sup>75</sup> A dánokat például már az 1970-es években sem érdekelte a hagyományos folklóranyag, számolt be Carsten Bregenhøj a dán folklórarchívum rövid történeti ismertetésében, és felvetette azt a kérdést, hogy vajon hogyan lehet az embereket érdekeltté tenni az intézmények fenntartásában, és hol lehet megtalálni a közönséggel az új kapcsolatot – *NIF Newsletter* 1978: 6(1). Magyarországon ugyan a társadalmi gyűjtőhálózat aktivitása éppen az 1960–70-es években érte el a csúcspontját (kb. 1200 pályázóval), ugyanakkor már abban az időben is parttalannak érezték a szakmabeliek, hiszen az aktuális tudományos célokhoz nem lehetett felhasználni azokat (Forrai 2000: 632).



A digitális textológia célja, hogy a számítógépes feldolgozás számára azonosítható módon jelölje a szöveghez tartozó minden olyan belső és külső információt, amely az elemzés szempontjából a későbbiek során releváns lehet. A digitális (szöveg)adat-előkészítés legnagyobb elméleti és módszertani paradoxona éppen ebben rejlik. Az adatot sokoldalúan kell előkészíteni, hogy minél többféle elemzésre alkalmas és minél több kontextusba visszahelyezhető legyen, az adatbázisok virtuális világában pedig minél több adathoz tudjon majd kapcsolódni, azokkal hálózatot alkotni. Ugyanakkor mégsem lehet egységesíteni az adat-előkészítést, mert az mindig attól függ, hogy mi az adott kutatás célja, amit ráadásul sokszor nehéz előre látni a digitális bölcsészettudományok esetében (Jockers és Underwood 2016: 299–300; McGann 2016).

A kiválasztott jelölőnyelvnek egyszerre kell tehát egységesnek, ugyanakkor mégis rugalmasan alakíthatónak és bővíthetőnek lennie. Ehhez jelenleg úgy tűnik, az egyik legjobb választás a szövegeket XML (Extensible Markup Language – kiterjeszhető jelölőnyelv) jelölőnyelvben kódolni, mely platformfüggetlen, és viszonylag ellenáll a technikai változásoknak. Maga a TEI (Text Encoding Initiative – szövegekódolási kezdeményezés) – mely a digitális tudományos szövegkiadások számára fogalmaz meg szövegekódolási javaslatokat – is az XML jelölőnyelvet ajánlja, ugyanis az nem véges elemkészletet határoz meg, hanem csupán rögzíti azokat a szabályokat, amelyek segítségével elkészíthető egy jól formázott jelölőnyelv.<sup>76</sup> A megfelelően kiválasztott jelölőnyelv azonban csak az érem egyik oldala – a következő problematikus kérdés, hogy mit és hogyan jelöljünk (McGann 2016).

A folklórisztikában a korábban már említett finn és észti runoadatbázisok, a román szerelmi ráolvasás-adatbázis az elsők között használták az XML jelölőnyelvet (Saarinen 2001), de az *ETKSpace* és a nemrég elindult ír *The Schools' Collection* projekt is ezt alkalmazza, ugyanakkor szövegekódolási eljárásaik a különböző célok miatt nem egységesek. Amennyiben használják is a TEI-t a folklórisztikai adatbázisok, azok elemkészletéből nem feltétlenül ugyanazokat az elemeket tartják jelölésre érdemesnek. Az *ETKSpace* egy hiedelem-szövegének XML kódolásából kiderül, hogy a készítők valójában mit kódoltak. Láthatjuk a metaadatként szöveghez rendelt kulcsszavakat („Keywords”), a szövegben szereplő helyneveket („Places Mentioned”), az egyes történetek Kristensen-féle tipológiai besorolását („ETK Index”), a szövegek azonosítóját („Story”), majd végül a szöveget („Story Text”):

6. ábra. Az *ETKSpace* egy hiedelem-szövegének XML jelölőnyelvben kódolt átirata

```
This XML file does not appear to have any style information associated with it. The document tree is shown below.
<?xml version="1.0" encoding="UTF-8" >
<root>
  <metadata type="Keywords"><wette/></metadata>
  <metadata type="Keywords"><ruustenen/></metadata>
  <metadata type="Places Mentioned"><odderup/></metadata>
  <metadata type="Story"><ID_82_0_0001/></metadata>
  <metadata type="ETK Index"><Hound dwellers (Hidden Folk)></metadata>
  <metadata type="Story Text">
    11. På en gård i Odder up er der sådan en ruustenen på ofæt med alle ting på et bestemt stiel og en bestemt tid ; men kun hove der er født på højtidseftene kan se, hvad det er ; og de siger, at det er nogle gamle små vætter. F. L. Grundtvig.
  </metadata>
</root>
```

Forrás: etk-space.scandinavian.ucla.edu<sup>77</sup>

<sup>76</sup> <http://www.tei-c.org/index.xml> (letöltve: 2017. október 4.). A TEI-ről lásd részletesen legutóbb Pierazzo (2016).

<sup>77</sup> <http://etk-space.scandinavian.ucla.edu/etkSpace/export/export.php?col=Story&ID=1&stylesheet=XML> (letöltve: 2017. szeptember 28.).

Ehhez képest az észtrunoadatbázis egy szövegének esetében jelölték a gyűjtés helyét, a gyűjtőről és adatközlőről tudott információkat (pl. hány éves), valamint a tényleges szöveget tartalmazó részben a sortöréseket is.

7. ábra. *Eesti regilaulude andmebaas* (The Database of Estonian Oral Poetry) egy runo szövegének XML jelölőnyelvben kódolt átírat

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet href="http://www.folklore.ee/regilaul/regifont.css?
20141027" type="text/css"?>
<!--?xml-stylesheet type="text/xsl" href="andm.xsl"-->
<KOGUM xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="XSDregilaul.xsd">
<ITEM nro="e0000000X0078">
<META>
<ID>E X 17 (78)</ID> &lt; <LOC>Elva al.</LOC> - <COL>M. J. Eisen</COL>
&lt; <INF><NIMI>Marie Kivi</NIMI>, <ELUL>65 a</ELUL></INF>
<TYP_KONT>tüübimised kontrollimata</TYP_KONT>
<TYP_YHT>Sirbiviskamise laul</TYP_YHT>
<LLIIK_YHT>Tõõlaulud </LLIIK_YHT>
</META>
<TEXT>
<V>Sirise, sirise, sirpi,</V>
<V>kõrise, kõrise rauda,</V>
<V>kelle sirpi ette jõuab,</V>
<V>selle juurde peigu jõuab,</V>
<V>Sellel peigu põlle ostab,</V>
<V>Ja tasuks tanugi ostab,</V>
<V>Passib pähä naise mütsi,</V>
<V>Naise mütsi, pruudi paelad.</V>
</TEXT></ITEM>
</KOGUM>
```

Forrás: folklore.ee<sup>78</sup>

A *The Schools' Collection* felületén minden egyes átírt szöveg esetében akár három XML fájl is letölthetünk. Egy vonatkozik az iskolára, egy az adott kézirati oldalra, egy pedig magára a szövegre. A három példa jól szemlélteti, hogy a textualizációs eljárások között azonos jelölőnyelvet választva is eltérő mélységű és részletességű szövegkiadások születnek. Még nagyobb lehet az eltérés, ha egy a folklorisztikában is fontos műfajt más diszciplína képviselői tesznek közzé. Az inkább irodalom- és könyvtörténeti megközelítésű *EBBA* (*English Broadside Ballad Archive*) is a TEI-jelöléseket alkalmazta. A ponyvaballadák kódolása azonban itt tartalmazza a kora újkori nyomtatvány precíz leírását (szerző, publikálás dátuma, kiadó stb.), az egyes balladák poétikai jellemzőit (stanzák, sorok, refrének), sőt a kiadás tipográfiai sajátosságait (kurzív, kapitális stb.) – ez utóbbi elvileg az objektum megsemmisülése esetén is pontosan visszaadná a nyomtatványt, hiszen minden létező materiális tulajdonságát kódolták benne.<sup>79</sup>

78 <http://www.folklore.ee/regilaul/andmebaas/?op=1&oid=4> (letöltve: 2017. szeptember 28.).

79 <http://ebba.english.ucsb.edu/page/tei-xml> (letöltve: 2017. szeptember 28.).

8a. ábra. Az EBBA (*English Broadside Ballad Archive*)  
egy ponyvaballadájának XML jelölőnyelvben kódolt átirata

```
</teiHeader>
<v:text>
  <v:body>
    <v:div type="ballad">
      <v:div type="part" n="1">
        <v:head>
          <v:title>
            <v:seg n="1" rend="left">
              <hi rend="italic">Witchcraft discovered and punished.</hi>
            </seg>
            <v:seg n="2" rend="left">
              <hi rend="italic">
                Or, the Tryals and Condemnation of three Notorious Nitches, who were Tryed
              </hi>
            </seg>
            <v:seg n="3" rend="left">
              <hi rend="italic">
                the last Assizes, holden at the Castle of
                <hi rend="bold">Exeter,</hi>
                in the County of
                <hi rend="bold">Devon:</hi>
                where they received Sentence
              </hi>
            </seg>
            <v:seg n="4" rend="left">
              <hi rend="italic">
                for Death, for bewitching several Persons, destroying Ships at Sea, and Cattel by Land, etc.
              </hi>
            </seg>
            <lb/>
            <v:seg n="5" rend="left">
              <hi rend="italic">
                To the Tune of,
                <hi rend="bold">D</hi>
                octor
                <hi rend="bold">Faustus:</hi>
                OR,
                <hi rend="bold">Fortune my Foe.</hi>
            </seg>
          </v:head>
        </v:div>
      </v:div>
    </v:body>
  </v:text>
```

Forrás: ebba.english.ucsb.edu<sup>80</sup>

8b. ábra. Az EBBA (*English Broadside Ballad Archive*)  
egy ponyvaballadájának XML jelölőnyelvben kódolt átirata

```

      <hi rend="italic">Stigian</hi>
      shore:
      </l>
      <l n="55" rend="left">For the great mischiefs she so oft had done,</l>
      <l n="56" rend="left">And wondered that her Life so long had run.</l>
      </lg>
      <lg>
      <l n="57" rend="left">She said the Devil came with her along,</l>
      <l n="58" rend="left">Through Crouds of People, and bid her be strong:</l>
      <l n="59" rend="left">And she no hand should have, but like a Lyer,</l>
      <l n="60" rend="left">At the Prison Door he fled, and nere came nigh her.
      </l>
      </lg>
      <lg>
      <l n="61" rend="left">The rest aloud, cravd Mercy for their Sins,</l>
      <l n="62" rend="left">Or else the great deceiver her Soul gains;</l>
      <l n="63" rend="left">For they had been lewd Livers many a day,</l>
      <l n="64" rend="left">And therefore did desire that all would Pray</l>
      </lg>
      <lg>
      <l n="65" rend="left">To God, to Pardon them, while thus they lie</l>
      <l n="66" rend="left">Condemned for their Wicked Deeds to Die:</l>
      <l n="67" rend="left">Which may each Christian do, that they may find</l>
      <l n="68" rend="left">Rest for their Souls, though wicked once Inclind.</l>
      </lg>
      </div>
      </closer></closer>
    </div>
    <v:closer>
      <v:seg n="1" rend="left">
        <hi rend="italic">FINIS.</hi>
      </seg>
    </closer>
  </div>
</body>
</text>
</TEI 2>
```

Forrás: ebba.english.ucsb.edu<sup>81</sup>

80 <https://ebba.english.ucsb.edu/ballad/31034/ebba-xml-31034> (letöltve: 2017. szeptember 28.).

81 <https://ebba.english.ucsb.edu/ballad/31034/ebba-xml-31034> (letöltve: 2017. szeptember 28.).

A digitális textológia tudományközi egységesítése nyilvánvalóan utópisztikus (rém)álom, az uniformizáció gyakorlatilag a szövegek és a különféle textualizációs eljárások következtében létrejött plurális olvasat lehetőségét semmisítené meg. A kiadásoknak elsődlegesen mindig az anyag természetéből kell kiindulniuk és bizonyos kérdésekre választ adniuk – még ha ezek a kérdések a digitális világban megsokszorozódnak is. Ugyanakkor olyan vélemény is van, hogy ha mindenki mást és mást annotál, a korpuszok elhatárolódnak egymástól. Martin Wynne, az oxfordi egyetem digitális bölcsészettudományi szakembere szerint a hoszszas és aprólékos adat-előkészítéssel a tulajdonképpeni adat értelmezése helyett a későbbi interpretációs lehetőségekre fókuszálunk, ezt pedig csak úgy küszöbölhetjük ki, ha az automatikus annotációhoz szükséges jobb és gyorsabb eszközök fejlesztésén dolgozunk inkább. Ebben az esetben a programok különböző digitális eljárások által saját maguk indexálják a korpuszokat (Wynne 2012).<sup>82</sup>

A fenti bizonytalanságok ellenére fontos, hogy a folklorisztika is kialakítsa a tudományának megfelelő metaadatrendszerét és annotálási technikákat, mert ha nem teszi ezt meg, elveszítheti a számára igazán fontos tulajdonságokat. A TEI például elsődlegesen irodalomtudományi-nyelvészeti indíttatású, így filológiai megközelítésű. Egy folklorisztikai kiadás esetében a bekezdések, paragrafusok jelölése azonban nem annyira mérvadó (hacsak nem tudománytörténeti vagy textológiai kutatást kívánunk készíteni), annak jelölése viszont, hogy az adott folklórszöveg címe vagy műfaji meghatározása émiikus vagy étikus; az adott lejegyzés emlékezetből, gyorsírással vagy hangrögzítésből történt; a gyűjtés szabad interjú, kérdőív vagy résztvevő megfigyelés eredménye, már annál indokoltabb lenne, még akkor is, ha sok esetben mindez egyáltalán nem tudható. Ráadásul a digitális szövegkiadások világában a korábbi *főszöveg* előállítását felváltotta a *szövegforrás* textualizációja.<sup>83</sup> A földrajztörténeti módszer(ek) katalógusrendszerei, a folklorarchívumok cédulái ennek még megfelelnek, azonban a későbbi, oral history jellegű folklóranyag, a terepjegyzetek és egyéb járulékos feljegyzések már nem objektivizálhatók és szegmentálhatók ilyen egyszerűen, a deskriptív, a fizikai objektumokat leíró metaadatsablonok nem kedveznek a szellemi örökség és az oralitás digitális archiválásának.

Az annotáció körül sok a vita, azonban az egyértelmű, hogy a különböző adatbázisok és korpuszok között a valódi átjárás akkor valósul(hat) meg, ha az adatbázis készítői és tervezői a szövegeknek ugyanazokat a tulajdonságait annotálják, és ezeket lehetőleg ugyanolyan rendszerben, vagy legalábbis egymással kompatibilis rendszerekben rögzítik.<sup>84</sup> Az összehangolhatóság és átjárhatóság az adatbázisok között kulcsfontosságú, amire igazán jó

---

82 A különböző állásfoglalásokról a kérdést illetően továbbá lásd Pierazzo írását (2016: 316–318).

83 Debreczeni Attila szerint a szövegforrás fizikai léttel bír, kéziratos vagy nyomtatott dokumentum. A főszöveg-szövegforrás váltásról lásd bővebben Debreczeni írását (2014: 32).

84 A digitális humán tudományokon belül számtalan kezdeményezés (pl.: Dublin Core Metadata Initiative; DCMI; Open Archives Initiative Protocol for Metadata Harvesting; OAI-PMH; CLARIN) van arra nézvést, hogy standardizálják ezeket a metaadatokat a csere reményében (Meder 2014a: 125). A különböző nemzeti ráolvasás-adatbázisok digitális textualizációs eljárásait egységesítő javaslatához vö. Ilyefalvi tanulmányát (2017).

folklorisztikai példát egyelőre nem lehet hozni.<sup>85</sup> A nemzetközi sztemderdek kifejlesztése a folklorisztikában még várat magára, habár ennek szükségességére az utóbbi időben többen is felhívták a figyelmet.<sup>86</sup>

*Módszertani és elméleti kérdések. Megbízhatóság, hitelesség.*

A jelölőnyelv és a jelölések kiválasztása azért ennyire kritikus és vitatott pontja a digitalizálási folyamatnak, mert nem pusztán technikai kérdés. Gyakorlatilag a fontos, a kiadást lényegesen meghatározó dolgok itt dőlnek el, hiszen a textualizáció határozza meg, hogy később mire lesz használható a digitális adat, utólagosan pedig nehéz korrigálni (McGann 2016; Pierazzo 2016; Debreczeni 2014: 28). Az adatbázis létrehozóinak tehát a megtervezés fázisában már tisztában kell lenniük azzal, hogy mire akarják alkalmazni azt, ugyanakkor a digitális adatbázisok az állandó bővíthetőség és kiegészíthetőség jegyében is készülnek. A folytonos megújíthatóság és instabilitás a korábbi tudományos gyakorlatokhoz képest bizonytalanságot szül. A könyvalapú szövegkritikai kiadások sem voltak soha tökéletesek, azonban létrejöttük pillanatában elméleti, módszertani és gyakorlati szempontból is lezárt alkotások voltak, s így kerültek be a tudományos kutatások világába (McGann 2014: 26). A digitális adatbázisok természetükből fakadóan mindezt soha nem fogják elérni. A megbízhatóság és hitelesség szempontjából azonban még ennél is sokkal zavaróbb, hogy a legtöbb esetben a digitális szövegi kiadások nem ismertetik azokat a textualizációs eljárásokat, melyeken szövegeik átestek, azaz nem fedik fel a szövegek „történetét” és nem teszik elérhetővé azokat nyers formájukban. Nem látatják a teljes korpuszt, nem derülnek ki a szövegválogatás pontos szempontjai, nem derül ki, hogy mi az, ami például nincs benne az adatbázisban. Ellen-

---

85 Az első ilyen projekt lehet az ISEBEL: *Intelligent Search Engine for Belief Legends*, mely keresőmotorja a holland, dán s a mecklenburgi folkloradatbázisok szövegein egyszerre végez majd elemzéseket. Theo Meder a göttingeni SIEF (2017) kongresszus egyik archívumokról szóló paneljében adott arról hírt, hogy pályázati forrásból lehetőségük lesz Timothy R. Tangherlinivel és Christoph Schmitt-tel együttműködve a három adatbázis összehangolására és egy keresőmotor kifejlesztésére. Lásd továbbá <https://www.researchgate.net/project/Trans-Atlantic-Digging-into-Data> (letöltve: 2017. június 20.). Az integrálhatóság szempontjából a folklorarchívumok esetében lényeges az is, hogyan tudnak más nemzeti/nemzetközi intézmények digitális archíválási standardjához csatlakozni és ennek milyen következményei, előnyei és hátrányai vannak. A jól átgondolt közös projektek nagymértékben segíthetik és egyszerűsíthetik a digitalizációt. Az észtl irodalmi múzeumban található észtl folklorarchívum fontos kéziratának szkennelése (például az alapító Jakob Hurt hagyatéka), a múzeum digitalizálási projektjében történik. A korábban bemutatott észtl runoadatbázisban tehát nem szükséges közölniük a kéziratok verziókat képfórmátumban, hiszen a múzeum adatbázisában nagyon hamar meg lehet jeleníteni azokat, s nyilvánvalóan csak idő kérdése, hogy mikor sikerül ezt egyenesen beépített linkekkel elérni, külön manuális keresés nélkül. Jó példa továbbá az intézmények és projektek közötti együttműködésre az izlandi hiedelemszöveg-adatbázis is. Bár a teljes szövegeket a *Sagragrunnur*-ban nem tudjuk elolvasni, viszont azoknál az eseteknél, ahol az adott kötetet a National and University Library of Iceland már digitalizálta, beépített linkkel azonnal a kötet hivatkozott oldalára navigálhatunk, s így a szöveget egészében elolvashatjuk; [www.Bækur.is](http://www.Bækur.is) (letöltve: 2017. június 21.).

86 Emellett a SIEF Working Group on Archives munkacsoportjának 2015-ös zágrábi munkaülésén felmerült a folklorisztikai adatbázisok metaadatainak rögzítését egységesítő vezérfonal terve is (Meder 2014a: 124; Katajamäki és Lukin 2013: 13). A SIEF Working Group on Archives célkitűzéseit lásd: <http://www.siefhome.org/wg/arch/> (letöltve: 2017. szeptember 29.).

példaként hozható fel szinte egyedüli, üdítő kivételként a holland meseadatbázis, amely a több mint 40 000 szöveget tartalmazó korpuszát a felhasználók számára is jól látható módon többféle szempont szerint számszerűsíti. A felület jobb oldali sávjában például a teljes (és egyébként folyamatosan bővülő) állományra vonatkozóan kapunk kimutatásokat, gyakorlatilag az összes fontosabb, az adatbázisban jelölt metaadatról. Ilyenek például a forrástípus, műfaj, téma, gyűjtő, gyűjtési idő/hely, nevek a szövegekben stb., amelyekre kattintva azonnal a kívánt szövegekhez navigálhatunk. A rendezés a leggyakoribbtól kezdve halad lefelé csökkenő sorrendben. Például a forrástípus tekintetében láthatjuk, hogy a „mondeling”, azaz a szóbeli gyűjtésből származó szövegből van a legtöbb, 26 115 adattal, a szövegekben szereplő nevek közül pedig Isten (God) vezet összesen 933 találattal. A holland meseadatbázisban a korpusz áttekintését a felhasználó által működtethető dinamikus idődiagramok is segítik.

9a. ábra. Részletek a *Nederlandse VolksverhalenBank* bal oldali sávjából: a forrástípusok számszerűsítése

Gronings (1396)	
Vlaams (1020)	
Noord-Brabants (927)	
[+] Toon resterende 44	
<b>Type bron</b>	
mondeling (26115)	
boek (8360)	
internet (3473)	
e-mail (1581)	
vragenlijst (1010)	
brief (908)	
[+] Toon resterende 17	
<b>Subgenre</b>	
sage (25724)	
mop (10460)	
broodjeaapverhaal (3024)	
raadsel (2125)	
sprookje (1921)	
personal narrative (1040)	
[+] Toon resterende 7	

9b. ábra. Részletek a *Nederlandse VolksverhalenBank* bal oldali sávjából: a szövegekben előforduló szavak számszerűsítése

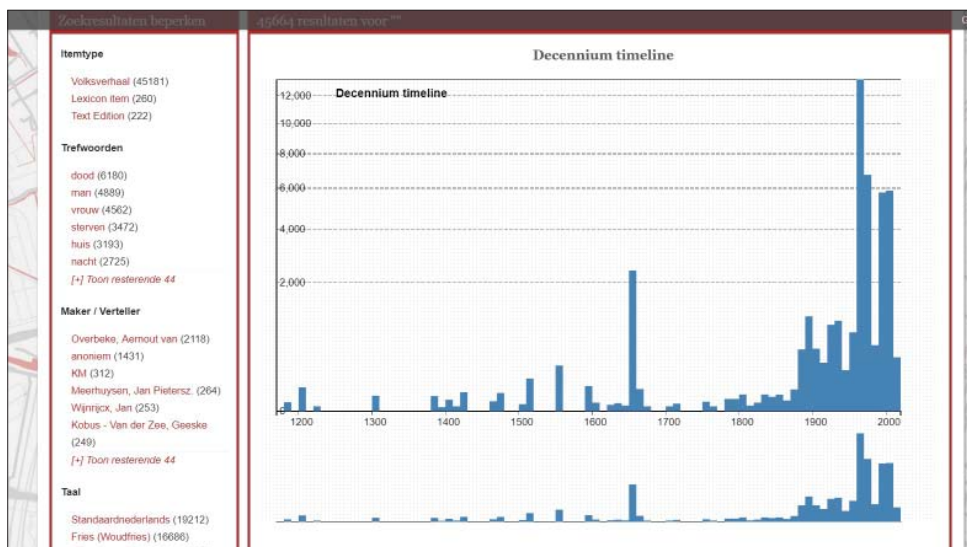
Plaats van Handelen	
Helmond (Noord-Brabant) (60)	
Broek in Waterland (Noord-Holland) (50)	
Schiedam (Zuid-Holland) (48)	
Oostermeer (46)	
Wittenberg (45)	
Amsterdam (Noord-Holland) (41)	
[+] Toon resterende 44	
<b>Naam Overig in Tekst</b>	
God (933)	
Sterke Hearke (623)	
Jan (417)	
Belg (403)	
Hearke (251)	
Nederlander (250)	
[+] Toon resterende 44	

Forrás: verhalenbank.nl<sup>87</sup>

87 <http://www.verhalenbank.nl/visuals/timeline?q=&facet=&free=> (letöltve: 2017. szeptember 28.).



9c. ábra. Részletek a *Nederlandse VolksverhalenBank* bal oldali sávjából: a meseadatbázis dinamikus időszávja



Forrás: verhalenbank.nl<sup>88</sup>

A fentiekkel ellentétben a *Sagnagrunnur* számszerűsített, áttekintő megjelenítést egyedül a kulcsszavaknál nyújt.

10. ábra. A *Sagnagrunnur* kulcsszókészletének számszerűsített áttekintése

The screenshot shows the 'Sagnagrunnur' interface with a table of topic words. The table has columns for 'id', 'Topic word', and 'Legends'.

id	Topic word	Legends
207	orðtoik (sayings)	107
19	prestar (priests)	1550
1024	pálsmessia (St Paul's Day)	1
85	páskar (easters)	56
288	púkar (demons)	48
527	refa- og minkaveiðar (fox and mink hunting)	52
615	refsing (punishments)	34
165	reimleikar (hauntings)	609
344	reki (driftwood)	20
488	reykjavik	56

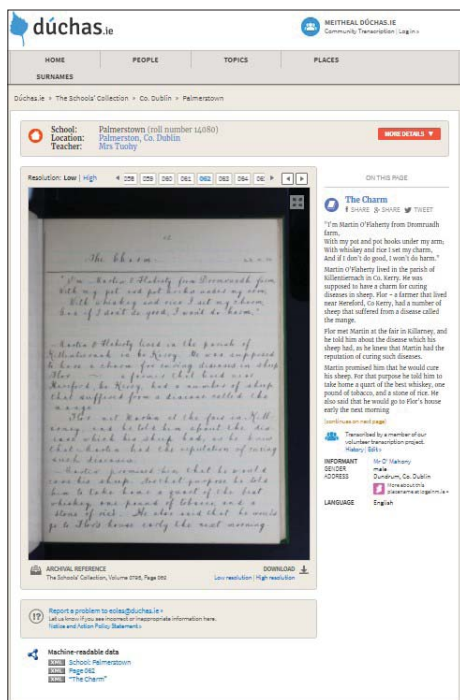
Forrás: sagnagrunnur.com<sup>89</sup>

88 <http://www.verhalenbank.nl/visuals/timeline?q=&facet=&free=> (letöltve: 2017. szeptember 28.).

89 <http://sagnagrunnur.com/grunnur/tags.php?lang=en> (letöltve: 2017. szeptember 28.).

Nem közli a szövegátírási elveket a *Danish Folklore Nexus* vagy a portugál legenda-adatbázis, és a példákat lehetne még sorolni. Sok esetben, ahogy fentebb már utaltam rá, az sem derül ki, hogy a monitoron látott szöveg pontosan honnan származik. Ennek persze pozitív ellenpéldái is ismertek, az ír *The Schools' Collection* projektben a kéziratosokat oldalanként látjuk beszkenelve, és az önkéntesek is oldalanként írhatják át azokat, valamint minden egyes oldal alján megtaláljuk azt is, hogyan kell az adott oldalra pontosan hivatkozni. Az önkéntes átírók számára átírási segédletet és bemutató videót is készítettek.<sup>90</sup>

11. ábra. A *The Schools' Collection* megjelenítési felülete



Forrás: [duchas.ie](http://duchas.ie)<sup>91</sup>

Néhány folklóradatbázisban, pl. a román szerelmi ráolvasások, a finn és észti runo, az *ETKSpace*, az ír *The School' Collection* hozzáférhetővé teszik az általuk használt és kódolt szöveg XML letöltését, ugyanakkor mások – a holland meseadatbázis, a pánhiszán ballada, a portugál legenda-adatbázis vagy a lett folklóarchívum adatbázisai – nem. Pedig a későbbi felhasználhatóság szempontjából ez rendkívül fontos lenne. A digitális bölcsészettudomány képviselői szerint a *nyers adatok (raw data)* közzététele és az adatbázisok önreflexív dokumentációja elengedhetetlen lesz a jövőben, a digitális szöveg folyton változó jellege miatt

90 <https://www.duchas.ie/en/info/meitheal> (letöltve: 2017. augusztus 16.).

91 <https://www.duchas.ie/en/cbes/4428221/4386886/4456663> (letöltve: 2017. szeptember 28.).

pedig az elemzéseket is időről időre le kell futtatni, és nem szabad azokat statikusan, egy állapotot tükröző végleges eredményként publikálni (Rieder és Röhle 2012: 81).

A tudományos digitális adatbázisok tartalmának előállítására nehéz, komplex feladat, ami új kihívások elé állítja a bölcsészeket. Ugyanakkor párhuzamosan az akkurátus és lelkiismeretes tudományos textológiai munka mellett a világban tömeges, főként kereskedelmi (pl. GoogleBooks) és könyvtártudományi indíttatású digitalizáció folyik, melyről szintén fontos szólni néhány szót, a világhálón elérhető digitális tartalmak nagy részét ugyanis pontosan e törekvéseknek köszönhetjük. A virtuális térben keringő *big data* tünékeny, filológiai értelemben megbízhatatlan és állandóan változó korpusza teljesen szemben áll a korábbi bölcsészettudományi gyakorlattal, ahol a kutatás első lépése a lehető legpontosabb és leg-hitelesebb szöveg előállítása volt. Kell-e foglalkozni ezekkel a szövegekkel, és ha igen, mire lehet őket használni? A vélemények megoszlanak ezzel kapcsolatosan is. Jerome McGann a kereskedelmi digitalizálás hátrányaira és a terület monopolizálásának (GoogleBooks) veszélyére hívja fel a figyelmet, kiemelve a megbízható digitális adat előállításának fontosságát (McGann 2014: 20–40). Jockers és Underwood szerint azonban a digitális paradigmában sok esetben csak az „elég jó” (*good enough*) szöveg szintjére lehet és kell törekedni, viszont bizonyos új típusú kutatások eleve nem igénylik a precíz szövegkiadást, tehát a „piszkos”, „zajos” szöveggel is lehet dolgozni (Jockers és Underwood 2016: 299–301). Leonard és Tangherlini a GoogleBookson megtalálható 19. századi dán regények korpuszán végzett elemzéséhez például teljesen elegendő volt ez a fajta szövegminőség is. Az *ETKSpace* szövegein egy szubkorpusz-témamodelláló algoritmussal meghatározták a fontosabb témákat, majd ezeket használva próbálták leszűrni további algoritmusokkal azt, hogy a GoogleBooksos korpusz mely 19. századi regényei szólnak a „vidékiességről”, a „rurális” életről. Ezzel lehetőségük nyílt megalkotni a további kutatás számára egy relevánsan előválogatott alkorpuszt (Tangherlini és Leonard 2013).

### Digitális módszerek és elemzések

A digitalizáció okozta anyagbőségben új módszerekre van szükség ahhoz, hogy az adatokat elemezzük, a közöttük lévő kapcsolatokat és dinamikát meglássuk. A digitális bölcsészettudományok sokféle módszert és eszközt használnak, köztük számos olyat is, mely csupán egy korábbi gyakorlat számítógépes változata (ilyenek például a statisztikai módszerek).<sup>92</sup> Bernhard Rieder és Theo Röhle digitális módszerekről írt kritikus összefoglalója szerint a számítógéphez mint új médiumhoz köthető módszerek közé tulajdonképpen csupán a *stimuláció*, az *adat felfedezése/felderítése* és az *automatikus vizualizáció (stimulation, data exploration, automated visualization)* tartozik (Rieder és Röhle 2012: 69–70). E módszerek közül a digitális bölcsészettudományban, így a számítógépes folklorisztikai kutatásokban is leginkább a különféle szövegbányászati, hálózatalméleti és vizualizációs módszerek és eszközök hoztak eddig sok újdonságot, ezért a következőkben ezeket mutatom be néhány példával.

---

<sup>92</sup> A digitális bölcsészettudományi módszerekről és azok fejlesztéséről lásd összefoglalóan legutóbb Hughes, Constantopoulos és Dalla (2016).

A számítógépes elemzések alapvető eszköztárához tartoznak a szövegbányászat (*text mining*) különböző módszerei. Amennyiben programok által *dekódolható szöveget* (*machine/computer-readable text*) állítunk elő, számtalan lehetőség áll az elemzések rendelkezésére az egyszerű szógyakoriság-statisztikáktól kezdve a kulcsszavazáson át a bonyolultabb témamodellálásig és látens szemantikai indexelésig.<sup>93</sup> A digitális adatbőség miatt a legtöbb esetben a legegyszerűbb, pusztán statisztikai elemzések is alkalmasak arra, hogy új megvilágításba helyezzenek szövegeket, korpuszokat. Ennél azonban sokkal többre képesek a szövegbányászati módszerek, amire jó példa a folklorisztika tekintetében a holland Meertens Institute keretén belül elkészített *MOMFER* (Meertens Online Motif FindER), azaz a Stith Thompson-féle *Motif-Index of Folk Literature* digitalizált verziója.<sup>94</sup> A motívumindexnek ugyan már létezett korábban is több online változata, viszont a *MOMFER* használta ki először a digitális módszerek nyújtotta többletlehetőséget. A többféle keresési opción és a gyorsaságán kívül (például nemcsak szavakra, hanem szókapcsolatokra is rá lehet keresni) a *MOMFER* újdonságát az adja, hogy a mutatót kibővítették egy szemantikus keresési alternatívával. A Thompson-féle index lemmatizált motívumaihoz hozzárendeltek egy angol szemantikus szótárt, a WordNetet, amely a szavak közötti többféle szintaktikai viszonyt is jelzi, szinonimakészlettel. A bővítés így lehetőséget ad általánosabb kategóriák feltárására és az egyes motívumok közötti összefüggések meglátására. A fejlesztők saját példái szerint, ha a színes állatok után keres bárki, akkor a zöld lovat is megtalálja, pedig se a zöldet, se a lovat nem adta meg keresési kulcsszóként, vagy a mérgezett gyümölcsökre rákeresve listázza a keresőmotor a mérgezett almát is. Ismerve a motívumindex tudományos értékelését a folklorisztika részéről (Dundes 1997), maguk a készítőik is megkérdőjelezik a *MOMFER* valódi hasznosságát, ugyanakkor egyszerű és gyorsan létrehozható mintapéldányát nyújtja egy korábbi folklorisztikai mű innovatív, digitális feldolgozásának (Karsdorp et al. 2015).

A szövegbányászati módszerek igazán nagy előrelépést jelenthetnek a folklorarchívumok és korpuszok soha véget nem érőnek tűnő katalogizálásában és rendszerezésében is. A manuális annotálás ugyanis rengeteg hibalehetőséget rejt magában, amelyek a különféle *vezérelt* (*supervised*) és *nem vezérelt* (*unsupervised*) automatizált technikákkal kiküszöbölhetők, ráadásul az anyagok mennyisége miatt is szükség van a tartalmak programok által végzett indexálására (Muiser, Theune és Meder 2012; Broadwell, Mimno és Tangherlini 2014). A lett folklorarchívum kampányában, ahogy fentebb említettem, engedélyezik a kulcsszó hozzáadását a begépelt szövegekhez, ami ugyan egy érdekes „folksonomy” kísérlet – tulajdonképpen arra, hogy a közreműködő laikus önkénteseknek mi jut eszükbe egy-egy 20. század elején gyűjtött folkloriszövegről –, de az igencsak megkérdőjelezhető, hogy a lett folklorarchívum ezáltal valóban rendezettebb lesz-e. A szubjektív egyéni kulcsszavazás archívumi munkatársak esetében is aggodalomra adhat okot (Granasztói 2008), viszont a kulcsszógenerálás újfajta lehetőségeivel igen gyümölcsöző kezdeményeket is láthatunk ma már. A Repository of Ireland (DRI) és a National Library of Ireland (NLI) közös vállalkozásaként hozták létre az ír folkloranyag kísérleti digitális tezaurusát (*MoTIF Pilot Thesaurus of Irish Folklore*), amit azóta a korábban említett *The Schools' Collection* anyagának tematikus rendszerezéséhez el

93 A szövegbányászati módszerekről lásd legutóbb Jockers és Underwood tanulmányát (2016).

94 <http://www.momfer.ml/>. (letöltve: 2017. április 10.).

is kezdtek használni a University College Dublin munkatársai. A digitális teaurusz megalkotásához az ír etnográfia és folklorisztika fontos kézikönyveit, folyóiratait és a nemzetközi folklorisztika egyes műveit használták fel, melyekből irányítottan nyerték ki és alkották meg a fogalomlistát (Ryan 2014a, 2014b). Amennyiben a tartalom teljes egészében olvasható számítógépek által (az ír esetben olvasható), akkor a programok automatikusan társítani tudják a szövegeket a különféle kategóriákhoz.

A szövegbányászat bonyolultabb eljárásainak segítségével további újszerű eredményt lehet elérni. A korábban említett, a 19. századi dán regények GoogleBooksos korpuszán végzett elemzés Tangherlini és Leonard szerint azért nagyon előremutató, mert az *ETKSpace* szövegeinek témamodellálásakor nem az volt a program feladata, hogy azonos vagy hasonló szövegeket találjon a korpuszon belül, hanem hogy „együttes szöelőforduláson alapuló szemantikus hasonlóságokat” mutasson fel. Éppen ezért szerintük a témamodellálás alkalmas arra, hogy felfedezzünk „témákat, amelyek azonos szemantikus »érzést« keltenek” (Tangherli és Leonard 2013: 741). A szövegbányászati eszközök – kiegészítve a különféle gráfelméleti és hálózatzelméleti módszerekkel – jól alkalmazhatók ott, ahol a kulcsszavas módszerrel nem lennének megtalálhatók az azonos szövegek. Kristensen saját maga idioszinkretikus tipológiájában például a fej nélküli lovasról szóló kísértettörténeteket az „anonim földesurak” (*unnamed manor lords*) kategóriába sorolta, Tangherlini és munkatársainak számítógépes elemzésével azonban sikerült a szöveget a kísértettörténetekhez sorolni, holott az nem tartalmazta a kísértet szót és annak szinonimáit sem (Abello, Broadwell és Tangherlini 2012). Egy másik elemzésükben a *Danske Sagn* szövegein a *látens témafelderítés* metódusát alkalmazták (LDA = *Latent Dirichlet Allocation*), melynek segítségével olyan témák emelkedtek ki a korpuszból, melyek előtte a Kristensen- vagy Tangherlini-féle klasszifikációban egyáltalán nem szerepeltek. Tangherlini szerint a további kutatás feladata azt megnézni, hogy ezek a szövegek vajon miért és hogyan tartoznak össze (2013a: 17–19).

A szövegbányászat több ponton is összeér a komplex hálózatkutatás módszereivel és eszközeivel. Jamshid J. Tehrani, a Durham University antropológusa a biológiában használt filogenetikai rendszert (törzsfajlódás során kialakult rokonsági kapcsolatok) és a kladisztika (leszármazási mintázatokat feltáró módszertan) elméletét és módszertanát felhasználva végző számítógépes elemzéseket többek között népmeséken (Tehrani 2013a). Első tanulmánya, melyet rendkívüli érdeklődés kísért a természettudományok képviselői részéről, megjelenése évében a száz legolvasottabb cikk között volt és több mint 73 000-en töltötték le – a *Piroska és a farkas* (ATU 333) és a *Farkas és a hét kecskegida* (ATU 123) mesetípusok közötti kapcsolatot vitatott folklorisztikai problémáját kívánta megoldani algoritmusokkal. Eredményeit éles kritika érte a CNRS francia folklorisztái részről, főként a vizsgált szöveganyagot (58 mese, ráadásul angol fordításban, a típusok gazdag német és francia adatait teljesen mellőzve), a fogalomhasználatot (mi a motívum) és a használt módszert (NeighborNet szoftver) illetően (Lajoie, d’Huy és Le Quellec 2013).<sup>95</sup> Azóta Tehrani sokféleképpen árnyalta az elemzés módszerét és a vizsgálati anyagot is, informatikusokkal együttműködve egy másik sokat vitatott folklorisztikai problémát kívánt újratárgyalni. A *Piroska és a farkas* vajon a szóbeliségből származik, azaz onnan adaptálta-e Perrault (majd ezt követően a Grimm fivérek) vagy az írásbeliségből? Tehrani és munkatársai az előbbit próbálják bizonyítani, s ezzel az utóbbi évek mesekutatásának állítását cáfolni (Tehrani, Nguyen és Roos 2015; Tehrani és

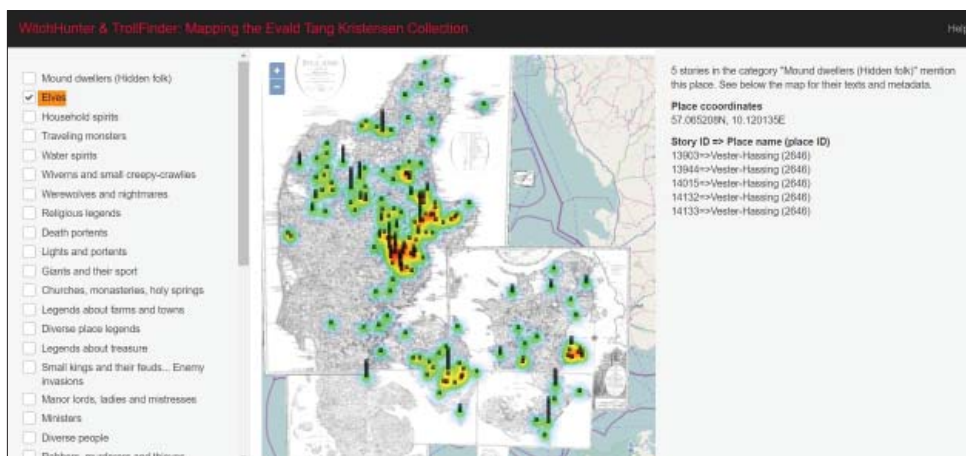
95 Tehrani válaszat a kritikára lásd Tehrani (2013b).

d’Huy 2017). Ugyancsak nagy médiaszenzációnak minősült Tehraninak a lisszaboni egyetem Institute for the Study of Literature and Tradition posztdoktorával, Sara Graça da Silvával írt közös cikke is, melyben számos jól ismert nemzetközi varázsmese korát határozták meg. *A kovács és az ördög* (ATU 330) meséről például azt állítják, hogy hatezer éves, azaz a szöveg gyökerei a bronzkorig vezethetők vissza (Graça da Silva és Tehrani 2016).

## Vizualizálás

A szövegbányászati módszerek mellett a legnépszerűbbek a különféle *vizualizációs eljárások*,<sup>96</sup> olyannyira, hogy egyesek szerint a digitális paradigmában egyenesen geográfiai és vizuális fordulatról kell beszélnünk (Presner és Shepard 2016). A vizualizációs eszközök közül az adatok digitális térképre vetítése emelendő ki elsődlegesen, ami a földrajzi információs rendszer (GIS = *geographic information system*) megjelenése óta terjedt el jelentősen. Dinamikus, a Googlemapsszel összekötött térképeket használ az izlandi hiedelem-szövegadatbázis vagy a holland meseadatbázis is. A digitális térképészet az adatok vizualizálásának módjait és lehetőségeit széleskörűen növeli. Az *ETKSpace*-ben Kristensen anyagát 19. századi, tehát a gyűjtés időpontjával nagyjából azonos történeti térképhez rendelték, illetve egy-egy téma sűrűségét ún. *forró térképek* (*heat maps*) segítségével jelenítik meg (Tangherlini 2013a: 22–23, 2016b; Tangherlini és Broadwell 2017).

12. ábra. *WitchHunter & TrollFinder: Mapping the Evald Tang Kristensen Collection*, az adatbázis *elf* adatainak forró térképes megjelenítése



Forrás: [etkspace.scandinavian.ucla.edu](http://etkspace.scandinavian.ucla.edu)<sup>97</sup>

A forró térképekre ráanyagíthatunk, az egyes függőleges oszlopok az adott településen gyűjtött szövegek számát jelzik (minél több, annál magasabbak). A digitális vizualizáció természetesen nem merül ki a térképekben. Jellemzően az adatok közötti multimodális és multidimenzionális kapcsolódási pontok, gráfok és hálózatok ábrázolására alkalmazzák még. Például a holland

<sup>96</sup> Összefoglalóan lásd legutóbb Sinclair és Rockwell (2016).

<sup>97</sup> <http://etkspace.scandinavian.ucla.edu/maps/witchhunter.html> (letöltve: 2017. szeptember 28.).



meseadatbázisban egy konkrét szöveg megjelenítésekor azonnal látjuk a jobb oldali sávban a szöveg lehetséges kapcsolódási pontjait, melyeket kattintással rögtön meg is nyithatunk.

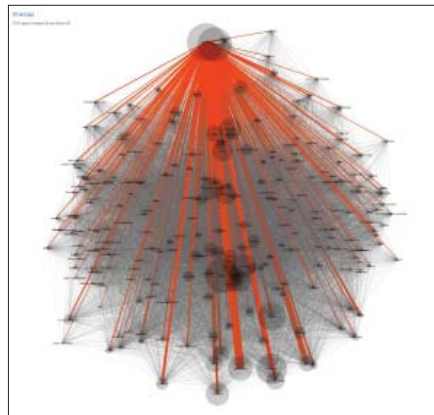
13. ábra. A *Nederlandse VolksverhalenBank* egy szövegének kapcsolódási pontjainak dinamikus vizualizációja a felület jobb oldali sávjában



Forrás: verhalenbank.nl<sup>98</sup>

Az adatvizualizációba ebben az esetben is belenyithat a felhasználó, melynek során a kapcsolódás pontok elhelyezkedése (minél távolabbi, annál kevesebb közös tulajdonsága van a két adatnak) és az összekötő vonalak vastagsága mentén (minél vastagabb, annál több metaadatuk egyezik) további értelmezésre nyílik lehetősége. Hasonló vizualizálási megoldást ajánl fel az izlandi *Sagnagrunnur* is a hiedelemszövegek kulcsszókészletét illetően.

14a. ábra. A *Sagnagrunnur* kulcsszavak közötti multidimenzionális kapcsolódási pontjainak vizualizációja

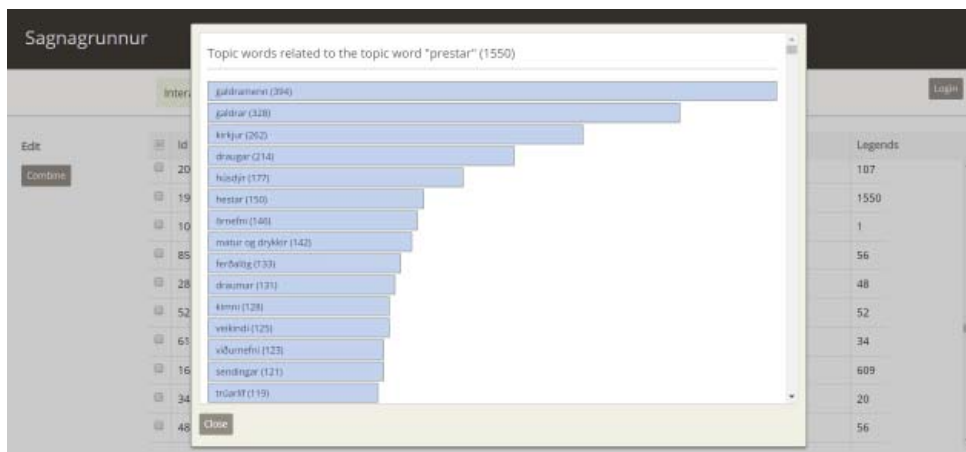


Forrás: sagnagrunnur.com<sup>99</sup>

98 <http://www.verhalenbank.nl/items/show/14643> (letöltve: 2017. szeptember 28.).

99 <http://sagnagrunnur.com/grunnur/tags.php?lang=en> (letöltve: 2017. szeptember 28.).

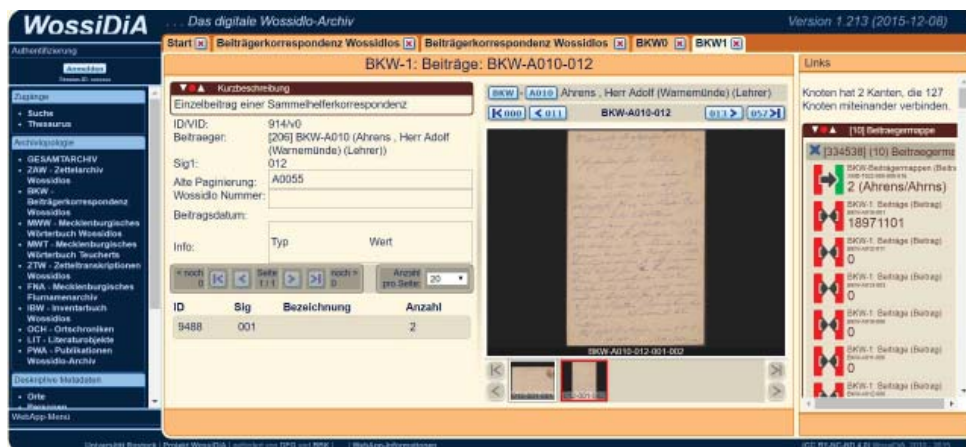
14b. ábra. A Sagnagrunnur kulcsszavak közötti multidimenzióális kapcsolódási pontjainak vizualizációja



Forrás: sagnagrunnur.com<sup>100</sup>

A WossDiA is számszerűsítve, valamint a fizikális archívumi fiókrendszert imitáló megjelenítéssel teszi láthatóvá a felület jobb oldalán az egyes adatokhoz köthető kapcsolódási pontokat.

15. ábra. A WossDiA egy adatának kapcsolódási pontjai



Forrás: apps.wossidia.de<sup>101</sup>

100 <http://sagnagrunnur.com/grunnur/tags.php?lang=en> (letöltve: 2017. szeptember 28.).

101 <https://apps.wossidia.de:8443/webapp/run> (letöltve: 2017. szeptember 28.).

Az adatvizualizáció természetéből fakadóan a leglátványosabb eleme a digitális bölcsészettudományi újításoknak. Ugyanakkor többen felhívják a figyelmet arra, hogy a színes, szemgyönyörködtető képeket kellő kritikával kell felhasználni, mert félrevezetők lehetnek, és – szemben egy argumentációval – nehezen cáfolhatók (tkp. egy új vizualizációra van szükség, mely mögött esetleg egy másik algoritmus áll; vö. Rieder és Röhle [2012: 73–75]). Nem csak azért idegen és gyanús tehát a bölcsészek egy részének a digitális bölcsészettudomány, mert ezernyi buktatót és kiszámíthatatlan problémát rejt az új szövegkiadási gyakorlat, hanem mert az elemzési folyamatokhoz gyakorlatilag nincs hozzáférésük. Az algoritmusok és gráfok bonyolult matematikai számításait a bölcsészek csak a legkritikábilis esetben képesek megérteni, hiszen nem erre képezték ki őket. Egy elemzésnél csak a bemenetet és kimenetet látják, a tulajdonképpeni, a program által véghezvitt elemzési műveleteket nem. Mivel a módszert, az elemzést magát nem tudják kritikálni, általában a kutatások egyéb elemeit (például a vizsgált korpusz minőségét és a kutatás elméleti alapvetéseit) kénytelenek. Tehrani és munkatársai fentebb említett állításait például nem lehet megcáfolni egy hagyományos bölcsészettudományi argumentációval, azokat csak egy másik vagy egy bővített korpuszon, vagy ugyanazon a korpuszon, de más algoritmusokkal és más szoftveres programmal végzett elemzés tudná igazolni vagy megdönteni. Ennek hiányában valójában csak arra mutathatunk rá, hogy a motívummeghatározás, amivel él, túlságosan szubjektív, vagy hogy az elemzés nem veszi számításba a különféle nyelvekből fordított szövegek fordításelméleti problémáit, stb. A *fekete doboznak* (*black-boxing*) is nevezett jelenség általános probléma a digitális bölcsészettudományon belül, megoldására a diszciplína teoretikusai az eddigieknél sokkal együttműködőbb, inter- és multidiszciplináris kutatásokat javasolnak, ahol a bölcsészek és az informatikusok folyamatos diskurzusban, valódi párbeszédben dolgoznak együtt (Rieder és Röhle 2012: 75–76; Lin 2012).<sup>102</sup>

### Elméleti keretek

Messze túlmutat e tanulmány keretein annak tárgyalása, hogy a digitális tartalmak és módszerek milyen nagyobb elméleti keretekbe és tudományos hipotézisekbe illeszkednek, mégis szükséges befejezésül röviden érinteni a témát, hiszen a digitális adatok és adatbázisok, valamint az elemzésükre használt digitális módszerek ugyanúgy súlyosan – ha nem súlyosabban – elméletekhez kötöttek, mint analóg társaik. Ezt azért fontos leszögezni, mert a technológiai innováció sokakat (s főként a tudományos eredmények populárisabb olvasóközönségét) egy olyan heurisztikus tévképzetbe lök(het), miszerint a számítógépes eljárások révén „végre” a bölcsészettudományi kutatások is objektívebbé, „mérhetővé”, e „kemény adatok” által pedig a természettudományokhoz hasonlóbbakká válnak, az adatvezérelt kutatások pedig végleg leszámolnak majd az elméletek és hipotézisek világával (Rieder és Röhle 2012; Schäfer és

---

102 A jelenséggel kapcsolatosan legutóbb Paßmann és Boersma arról értekeztek, hogy a kutatások során mindig is voltak és lesznek olyan „fekete dobozok”, melyeket nem lehet és nem is kell feltétlenül felnyitni. A fekete dobozok megszüntetése helyette fontosabb az, hogy egy kutató olyan kompetenciát fejlesszen ki, hogy bár nem érti az algoritmus működését teljesen, ennek ellenére azonban el tudja dönteni, mikor bízhat meg egy eredményben, és mikor kell esetleg megkérdőjeleznie azt (Paßmann és Boersma 2017: 141–142).

van Es 2017: 15).<sup>103</sup> Kevesen reflektálnak (a folklóradatbázisok készítői közül is alig néhányan) arra, hogy a digitális tartalmak, módszerek elméletileg már önmagukban mennyire terheltek. Csupán az önreflexív, második és harmadik hullámú digitális bölcsészettudományokban jelenik meg hangsúlyosabban a digitális tartalmak és módszerek episztemológiai sajátosságainak tudatosítása. Lássuk tehát, milyen elméleti keretekbe illeszkednek, milyen alapfeltevésekből indulnak ki a digitális bölcsészettudományi kutatások, mennyiben paradigmaváltók ezek, és a számítógépes folklorisztika hogyan viszonyul hozzájuk?

Az egyik legfontosabb elméleti és módszertani kiindulópont kétségkívül a *léptékváltás*, melynek megragadására sokféle kifejezéssel élnek a szerzők. Az általánosságban megfogalmazott *nagy léptékű* (*large scale*) adatelemzések, vagy a főként irodalmárok által használt terminus, a *távoli olvasat* (*distant reading*, vö. Moretti 2007), az irodalom *makroelemzése* (*macroanalyses*, vö. Jockers 2013) vagy a *folklorisztika makroszkópja* (*folklore macroscope*, vö. Tangherlini 2013a) megnevezések mind ugyanazt a (főként mennyiségi) perspektívaváltást próbálják érzékeltetni, azt a *madártávlati* (*bird's-eye view*) vagy megint mások szerint *isten* nézőpontot (*God's eye*) megragadni, ahonnan végre „láthatjuk a fától az erdőt” (Abello, Broadwell és Tangherlini 2012: 70). A szemlélet előfeltevése, hogy a korábbi egyes szövegek/jelenségek/adatok *közeli olvasatával* (*close reading*) ellentétben az adatok összessége által az egyes szövegek kontextusát, ebből kifolyólag pedig magukat az egyes szövegeket is jobban megérthetjük (Jockers 2013: 27). Többen megfogalmazták már, hogy e léptékváltás valójában nem a digitális bölcsészettudomány hozadéka, a kvantitatív bölcsészettudományoknak hosszú története van, gondolhatunk akár a 19. századi pozitivistádatgyűjtésre, az enciklopédikus szemléletre, vagy az adatokból kirajzolódó makrostruktúrák Annales-iskolához

---

103 Jó példa erre Tehrani eredményeinek magyar online médiareprezentációja. Bár a legtöbb esetben a folklorisztikai eredmények nem kimondottan címlaptörténetek, Tehrani kutatásairól már első komolyabb, a mesekutatást érintő publikációja előtt (Tehrani 2013b) a *The Daily Telegraph* közvetítésével hírt adott a *múlt-kor* elnevezésű tudománynépszerűsítő portál 2009 szeptemberében, a *Piros és farkas* népmesetípus orális eredetére vonatkozóan: [http://mult-kor.hu/20090911\\_tobb\\_ezer\\_evnyi\\_eroszakrol\\_szol\\_a\\_piroska\\_es\\_a\\_farkas](http://mult-kor.hu/20090911_tobb_ezer_evnyi_eroszakrol_szol_a_piroska_es_a_farkas) (letöltve: 2017. szeptember 28.). A hír ugyanabban a hónapban megjelent az alábbi portálokon is: <http://www.kultura.hu/tobb-ezer-evnyi>; <http://www.urbanlegends.hu/2009/09/piroska-es-a-farkas/> (letöltve: 2017. szeptember 28.). Tehrani négy évvel később, 2013. november 13-án publikálta hipotéziseit a *PLOS One* tudományos online folyóiratban, lásd <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078871> (letöltve: 2017. szeptember 28.). Egy nappal a megjelenést követően, 2013. november 14-én már MTI-hírként került az online hírek körforgásába, a magyar hírportálok közül számosan átvették és kisebb-nagyobb változtatással közölték azt. Például a *múlt-kor* portál: [http://mult-kor.hu/20131114\\_megfejtettek\\_a\\_piroska\\_es\\_a\\_farkas\\_eredetenek\\_titkat?fb\\_comment\\_id=319486061526857\\_1462200#f3e5958d011ed6c](http://mult-kor.hu/20131114_megfejtettek_a_piroska_es_a_farkas_eredetenek_titkat?fb_comment_id=319486061526857_1462200#f3e5958d011ed6c); a *vs.hu*: <https://vs.hu/kozelet/osszes/kiszamolnak-honnan-ered-a-piroska-es-a-farkas-1114>; az *Index* tudomány rovata: [http://index.hu/tudomany/2013/11/14/kiderult\\_honnan\\_ered\\_a\\_piroska\\_es\\_a\\_farkas/](http://index.hu/tudomany/2013/11/14/kiderult_honnan_ered_a_piroska_es_a_farkas/); a *Magyar Nemzet Online* felülete: <https://mno.hu/grund/megvan-a-piroska-es-a-farkas-eredetenek-titka-1195307>; a *csaladhalo.hu*: <http://csaladhalo.hu/szinek/megvan-a-piroska-es-a-farkas-eredetenek-titka/>; a *twice.hu* bulvár rovata: [http://twice.hu/bulvar/megfejtettek-a-piroska-es-a-farkas-eredetet; sőt az informatikai eszközöket árusító m.ipon.hu is: https://m.ipon.hu/hir/piroska-evolucioja/27724, vagy a szeretlekmagyarorszag.hu: http://www.szeretlekmagyarorszag.hu/honnan-ered-a-piroska-es-a-farkas-tortenete/](http://twice.hu/bulvar/megfejtettek-a-piroska-es-a-farkas-eredetet; sot az informatikai eszkozoket arusito m.ipon.hu is: https://m.ipon.hu/hir/piroska-evolucioja/27724, vagy a szeretlekmagyarorszag.hu: http://www.szeretlekmagyarorszag.hu/honnan-ered-a-piroska-es-a-farkas-tortenete/) (letöltve: 2017. szeptember 28.). Ugyancsak futótűzként terjedt Graça da Silvával 2016. január 20-án közösen publikált cikke (Graça da Silva és Tehrani 2016). Szintén a publikáció másnapján MTI-hírként, 2016. január 21-én négy, a januári-februári hónapban pedig további négy magyar hírportál hozta le a cikket. A hírekben az alábbi kifejezések dominálnak: „megfejtették”, „kiszámolták”, „megvan a titok”, „sikerült megfejteni” stb. A kutatási eredményeket értékelve egyedül a *National Geographic Magyarország* szolgáltató meg népmesekutató szakértőt. Bordás Veronika 2016. április 27-én megjelent „Mesekincsünk eredete” című írásában Gulyás Judit véleményét kérte (MTA Néprajztudományi Intézet, tudományos főmunkatárs), aki kritikai észrevételeinek adott hangot. Lásd <http://www.ng.hu/Civilizacio/2016/04/27/Mesekincsunk-eredete> (letöltve: 2017. szeptember 28.).

köthető *longue duree* koncepciójára is (Jockers 2013: 19; Jockers és Underwood 2016: 292; Kokas 2016: 407). Kicsit távolabbról a folklorisztika hagyományos földrajztörténeti irányzata is hasonló alapokról indult, hiszen a minél több variáns, a minél több szöveg összegyűjtése alapfeltétele volt annak, hogy egy-egy motívum, forma terjedési ívét és történetét megláthassák és felfejthessék, hogy végül ezekből a szöveg eredetét is meghatározzák. Timothy R. Tangherlini szerint éppen ezért a digitális folkloradatbázisokkal és számítógépes elemzésekkel egy „új földrajztörténeti irányzat” kialakulását láthatjuk, ahol azonban nem az eredet, az *Urform* megtalálása a végcél, hanem sokkal inkább a szövegek látens geoszemantikus viszonyainak megértése (Tangherlini 2013a: 21).

Miért lehet fontos a folklorisztikában mindez? Milyen többletjelentése lehet 30 000 hiedelemszövegnek madártávlatból, amit a korábbi, közeli olvasattal nem lehetett elérni? Az *ETKSpace* kutatói a *WitchHunter* & *TrollFinder* fantázianevet viselő platform segítségével (melyben többféle digitális eszközt és módszert integráltak) a szövegek és azok földrajzi elhelyezkedése között például azt állapították meg, hogy az óriásokról szóló történeteket kizárólagosan a valamikori nagy gleccservonulatok nyomvonalán mesélték, vagy hogy a főbb kereskedő útvonalak mentén jelentős mennyiségű történet maradt fenn a *buckalakókról* (*hidden folk/mound dwellers*), akik a dán folklór szerint jellemző módon kirabolhatják az úton lévőket, illetve hogy az *elfeket* következetesen több száz kilométer távolságba, a 19. századi Dániában még megműveletlen és vad Jutland területére helyezték el a mesélők (Tangherlini és Broadwell 2017: 144–151). A *GhostScope* & *TreasureX: Conceptual Geographies* elnevezésű felülettel az egy-egy adatközlő és az elmesélt történetek közötti földrajzi irányra kérdeztek rá, az elemzés kiindulópontja, hogy minden egyes történetmondó egy-egy konceptuális center, a földrajzi helyeket, irányokat és távolságokat ebből a pontból is érdemes megvizsgálni több ezer mesélő esetén. A vizsgálattal kimutatható volt, hogy a 19. századi Dániában a nők történeteiben zömében csak közvetlen lakóhelyük, maximum a szomszéd település jelenik meg, ezzel ellentétben a mobilisabb életmódot folytató férfiak narratívái távolabbi körzeteket és a fontosabb gazdasági pontokat is érintették (Tangherlini 2013a: 20). Az életvilágok, életmódok és a történetek földrajzának viszonyából az is kiderült, hogy míg az adatközlők általában a lakóhelyükhöz közeli eseményekről mesélnek, a hajósok esetében ez akár többszöröseire is elnyúlhat a part mentén (Tangherlini és Broadwell 2017: 139–140). Bár a fentiek első ránézésre nem rendkívüli eredmények, az adatbázis finomhangolásával, új kérdések megfogalmazásával annak a problémának a megértéséhez kerülünk közelebb, hogy az ember történetei által hogyan viszonyul a környezetéhez.

A nagyléptékűség a digitális bölcsészettudományokban a mennyiséget illetően nemcsak horizontálisan jelenik meg, hanem mélységében, vertikálisan is. Másik kulcsfogalma ugyanis az adatbázisoknak a *sűrű/tömött/mély korpuszok és térképek* (*thick/deep corpuses/maps*). A fogalomhasználatból kitűnik, hogy a szemlélet a geertzi sűrű leíráshoz kapcsolódik, igaz, jócskán a saját képére formálva azt (Presner et. al 2014: 18–19). Itt ugyanis gyakorlatilag a sűrűség annyit jelent, hogy a különféle adatsorok összekapcsolásával más-más perspektívából válik elemezhetővé ugyanaz az anyag, és hogy a kontextus által meghatározott jelentéskötésre helyezik a hangsúlyt. A szemlélet szerint az adatbázisok soha nem befejezettek, így mindig viszonylagosok, ugyanakkor nemcsak abból adódik „sűrűségük”, hogy több adatsort ábrázolnak egyszerre, hanem abból is, hogy ezáltal a megértés különböző útjait mutatják meg. A sűrű térképek egyes természettudományi diszciplínákat, pl. a korábbi leíró, statikus földrajztudományt ily módon a bölcsészettudományok megérté-

si módozataihoz, a jelenségek plurális értelmezéséhez közelítik (Presner et. al 2014: 18–19; Presner és Shepard 2016).

Az adatok, szövegek tömeges digitalizációja, a nagy léptékű elemzések számos esetben felülírhatják az eddigi, például irodalmi vagy folklórisztikai kánonokat, és soha nem látott módon tágitják a vizsgálati korpuszokat. Ugyanakkor nem árt tudatában lenni annak, hogy jelen pillanatban a digitalizáció sok esetben esetleges, és nyilvánvalóan nem csak az létezik, ami az interneten van, azaz, amit már digitalizáltak.<sup>104</sup> Az adatbázisok sikeres összehangolása után éppen ezért megfelelő kritikával kell kezelni a számítógépes elemzések eredményeit is. A folklórisztika tekintetében például számtalan csonkán maradt kezdeményezést, projektet találhatunk az interneten, az adatbázisok pedig nemcsak minőség, hanem a bevitt folklórszöveg mennyisége szerint is igencsak eltérnek egymástól. A román szerelmi ráolvasás-adatbázisban például mindössze 119 szöveget találunk, de ennél is szemléletesebb, hogy míg a belga vagy holland meseadatbázisok több mint 40 000 szöveget tartalmaznak, addig a katalán adatbázis kb. 6000 szövegére csak hivatkozásokat kapunk, a francia meseadatbázisban 2014-ben pedig 100 körüli szöveget helyeztek el (Meder 2014a). Vajon hogyan lehet súlyozni ezeket az aránytalanságokat a számítógépes elemzésekben?

A nagy léptékű adatelemzéseket végző kutatók számos helyen felhívják a figyelmet arra is, hogy az új, számítógépes lehetőségek nem váltják fel a korábbi bölcsészettudományi kutatásokat. A közeli olvasatot a távoli olvasat nem szünteti meg, mindkettőre szüksége van, sőt a kettő együttes alkalmazása, azaz a kettő közötti navigáció lehetősége az, ami igazán megújíthatja a kutatásokat (Abello, Broadwell és Tangherlini 2012; Jockers 2013).

A nagyléptékűség, a sűrű korpuszok, de a tanulmányban sokszor felmerülő hálózatelmélet is már jóval a digitális technológia megjelenése előtt a bölcsészettudományi gondolkodás része volt, igaz, az új médium rendkívüli mértékben segíti kiteljesedésüket. Már a hagyományos bölcsészettudományokban is az volt a legfőbb szerepük, hogy olyan struktúrákra, keretekre, jelentésekre mutassanak rá, melyeket egyébként nem vettünk volna észre. A *keret (pattern, frame)* a digitális bölcsészettudomány kedvelt, sokat használt bűvös terminus technicusa, annak ellenére, hogy meghatározása tudományfilozófiai problémákat vet fel, és az inkonzisztens fogalomhasználat miatt egyáltalán nem egyértelmű, hogy a hagyományos és digitális bölcsészettudomány mit ért alatta, sőt sokszor az sem, hogy mire alkalmazható (Rieder és Röhle 2012; Dixon 2012). Összevetve azonban az analóg kutatások keretfelismerését a digitális bölcsészettudomány eljárásaival, az a nagy különbség, hogy a kereteket ez utóbbiakban már a számítógépes programok ajánlják fel és nem a kutató veszi azokat észre.<sup>105</sup> Bár némely esetben egyelőre olyan helyeken is felkínálhatnak kereteket a számítógépes algoritmusok, ahol teljesen értelmetlenek és irrelevánsak (Tangherlini 2013a: 23), a digitális bölcsészettudomány paradigmaváltó nóvuma valójában ez. Ezen a ponton tulajdonképpen a *mesterségesintelligencia-kutatásokhoz* érkezünk el, ahol a cél, hogy a programokat ne csak olvasni tanítsuk meg, hanem arra is, hogy „értsék”, amit olvasnak. A tanulmány során érintett digitális módszerek és eszközök működése mögött a mesterséges intelligencia területéhez tartozó alap kutatások állnak (nyelvtechnológia, számítógépes nyelvészet, számítógépes narratológia). A szövegbányászat és a *természetes nyelvfeldolgozás* (NLP = *natural language process*) gyakorlatilag ehhez használnak különféle módszereket és eszközöket. Mindegyik-

104 A korpuszok egyenetlenségének problémájához vö. Jockers és Underwood tanulmányát (2016: 301).

105 A statisztikákon alapuló számítógépes keretfelismerő eljárásokról lásd Golden írását (2015).



nek szüksége van hozzá egy *konzisztens tudásbázisra* (*consistent knowledge base*), például részletes teauruszra, lexikonra, szabályos grammatikákra. A folklórszövegek kiváló homogén alpanyagot szolgáltatnak ehhez, éppen ezért nem meglepő, hogy a 2012-ben megrendezett *The Third Workshop on Computational Models of Narrative* konferencián az előadások nagy része folklóranyagon végzett kutatási eredményekről számolt be, a tanácskozás szervezője pedig az az informatikus, mesterségesintelligencia-kutató Mark A. Finlayson volt, aki doktori disszertációját (Massachusetts Institute of Technology) a proppi mesemorfológia számítógépes elsajátításából írta (Finlayson 2016).

## Konklúzió

A folklorisztika a kezdetektől szinte a mai napig küzd azért, hogy önállóan létező tudományként legitimálja magát. Katona Lajos, a századforduló (19–20. század) magyar folklorisztikájának egyik fontos alakja szerint az minősül önálló diszciplínának, melynek önálló tárgya és módszere van, a megszerzett ismeretek pedig a társtudományoktól elkülöníthetően önálló rendszert alkotnak (Landgraf 2016: 507). A 20. század eleji elgondoláshoz képest a posztmodernitás inter- és multidiszciplináris kutatásai jelentősen átalakították e tudományfelfogást, a különféle tudományágak határai ma már kevésbé élesek, a nagyfokú specializálódás következtében pedig számtalan új, átmeneti forma is létesült. Ráadásul úgy tűnik, a digitális bölcsészettudomány az akadémiai élet szervezeti és strukturális létmódját is gyökeresen átalakítja (Berry 2012; Evans és Rees 2012; Thomas 2016; McCarty 2016), miközben a tudás-termelés új formái jelennek meg. Még nem tudjuk, milyen lesz az a szövegkiadási gyakorlat, mely teljesen független a könyvalapú gondolkodástól, és milyen az a tudás-termelés, mely nem könyvekben és lábjegyzetekben gondolkozik – ezt csak a következő nemzedék látja majd, amely ebbe születik bele (Debreczeni 2014: 38–39).

A digitális bölcsészettudomány azonban még a 20. század eleji, hagyományosabb tudománykoncepcióknak megfelelően is egyre inkább önálló diszciplínaként működik.<sup>106</sup> Ha jobban szemügyre vesszük a számítógépes folklorisztika célkitűzései közé felsoroltakat, láthatjuk, hogy a legtöbb megegyezik a digitális bölcsészettudomány általános célkitűzéseivel és területeivel (Abello, Broadwell és Tangherlini 2012; Tangherlini 2014, 2016a). A digitalizáció gyakorlati oldala, a szövegek klasszifikációja, a kánonok és a korpuszok problémái, a digitális tartalmakon használat eszközök és módszerek (szövegbányászat, vizualizáció) és az ezek mögött meghúzódó elméletek (léptékváltás, sűrű korpuszok, keretfelismerés) egyáltalán nem kizárólagosan a folklorisztika problémái. Akkor mégis mi a folklorisztika a számítógépes folklorisztikában? Hol és pontosan mi a folklorista szerepe?

A sikeres digitalizációhoz egyrészt elengedhetetlenül szükségesek a terület szakértői, informatikai szakzsargonnal élve a *domain expert*-ek. Azaz, a tanulmány első felében bemutatott folklorisztikai és folklórarchívumi háttérismeretek és tudás nélkül nehezen képzelhető el egy jól működő tudományos folklóradatbázis felépítése és működtetése. Másrészt csak az tud megtervezni és kivitelezni egy folklóradatbázist, aki nemcsak az anyagot ismeri, hanem az új technika nyújtotta lehetőségeket és buktatókat is. Mindez azonban nem azt jelenti, hogy

---

106 A digitális bölcsészettudomány fő fejlesztési területei: 1. digitális tartalom, 2. digitális eszközök, 3. digitális módszerek. Vö. Hughes, Constantopoulos és Dalla (2016: 151).

a folkloristáknak vagy más bölcsészettudományi szakembereknek át kellene képezniük magukat matematikussá, informatikussá, és hogy a jövőben programot írni tudó folkloristákra lesz csupán szükség. Azt viszont igen, hogy a digitalizálási hullámban és lázban a folkloristák is átlássák a tudástermelés folyamatát, és képesek legyenek a maguk hasznára fordítani az új technológiai eszközöket. Fontos tehát, hogy a bölcsészek a 21. században is értsenek a tartalom előállításához, azért, hogy azokat megfelelő forráskritikával tudják kezelni, hiszen a számítógépes programok által generált és lehívott adatokat ugyanolyan kritikus és szigorú tekintettel kell egy kutatónak felhasználnia és értelmeznie, mint amikor az archívumból/levéltárból a történész előbányász valamit. Az adat sohasem passzív, az adat-előállítás aktív folyamat a digitális paradigmában is (Rieder és Röhle 2012; Jockers 2013; Kokas 2016: 412; Schäfer és van Es 2017).

A 2000-es évek legelején még úgy tűnt, a folkloradatbázis csupán egy eszköz, Voigt Vilmos szavaival élve ezért „örül neki mindenki”, és ezért nem lehet „cáfolni” vagy igazán kritizálni azt, mert pusztán egy új technikai innovációról van szó (Voigt 2006). Bár az adatbázis valóban egy eszköz, mögötte – ahogy az analóg szövegkiadás mögött is – számos elméleti és metodológiai döntés húzódik meg, melyekkel tisztában kell lenni ahhoz, hogy az eredményeket tudományosan értékelhető módon értelmezzük. Az adatvizualizáció vagy a szövegbányászat által felmutatott gócpontok és keretek nem a kutatás végcéljai, ahogy maga az adatbázis sem az. Az adatbázisok mellérendelő, enciklopédikus logikája, a keretek és gócpontok mit sem érnek a kutató értelmező narratívája nélkül, hiszen „a tények és a tények interpretációja között mindig lesz egy emberi mozdulat” (Jockers 2013: 30; vö. továbbá: Manovich 2009 [2001]; Kokas 2016: 412).

A megfelelő forráskritika elsajátításán túl a folkloristák legfőbb feladata tehát ott van, hogy a számítógép adta lehetőségeket ismerve folklorisztikai szempontból releváns kérdéseket tegyenek fel a digitalizált folkloranyagnak, főként olyanokat, amelyekre a válaszadás a korábbi technológia miatt lehetetlen lett volna (Meder et al. 2016: 93). Túlságosan merész lenne annak körülhatárolása, hogy mi számít ma folklorisztikai kérdésnek. Tangherlini tág, kicsit leegyszerűsített definíciója szerint a folklorisztika kezdetektől máig tulajdonképpen *emberek* (mesélők és kutatók), *helyek* (hol gyűjtötték és milyen helyekről tesznek benne említést) és *történetek* (vagy általánosságban véve a folklór kifejezőmódja) változatos viszonyának vizsgálatából áll (Abello, Broadwell és Tangherlini 2012: 65). Bár egyelőre kevés jó példa van arra, hogyan kell folklorarchívumot/gyűjteményt úgy digitalizálni, hogy az majd valóban a folklorisztikai kutatás eszközévé váljon (Tangherlini és Broadwell 2014: 225), a tanulmányban röviden összefoglalt kezdetleges kísérletek azt mutatják, hogy a digitális paradigmában a kutatók az összes eddigi fontos folklorisztikai problémát (textualizációs dilemmák, klasszifikáció és tipológia kérdése, oikotípus, strukturalizmus, terjedés és variálódás, szellemi kulturális örökség, eredet stb.) érintik, s a kérdéseket újra felteszik. Az eszközök és módszerek informatizálódásával pedig újra erősebben jelenik meg egy-egy természettudományi kutatási elmélet vagy modell (valószínűség-számítás és statikus fizika, filogenetikus biológia stb.) alkalmazása. Néhány szerző azonban óvatosságra int ez ügyben, miszerint megfontolandó, hogy biztosan szükséges-e a természettudományi elméletek átvetele, hiszen a bölcsészettudományok alapvetően másként működnek, céljuk nem feltétlenül a bizonyítás, hanem „kérdések fejlesztése és új meglátások felfedezése” (Ramsay 2003: 173, hivatkozva Schäfer és van Es 2017: 15).

A számítógépes folklorisztikai elemzéseknek igazi kritikája még nincs, művelői óvatosan fogalmaznak, egyelőre inkább eszközök és módszerek fejlesztése és próbája zajlik, merőben új kérdéseket és eredményeket még nem tudnak a művelői felmutatni. A tanulmányban vázolt számtalan dilemma és megválaszolatlan kérdés azonban nem a digitális textológia és főleg nem a folkloradatbázisok készítése vagy azok számítógépes elemzése ellen érvel, éppen ellenkezőleg. Bár a digitalizálás számos problémát és buktatót rejt magában, a kérdés továbbra is nyitott: hogyan érdemes tárolni, elemezni és mit lehet kezdeni a gigantikus méretűre duzzadt folklorarchívumokkal? A digitális bölcsészettudomány és azon belül a számítógépes folklorisztika egy új, de nem kizárólagos lehetőséget ad ezek megválaszolására.

## Hivatkozott irodalom

- Abello, James, Peter Broadwell és Timothy R. Tangherlini (2012): Computational Folkloristics. *Communications of the ACM* 55(7): 60–70.
- Anttonen, Pertti (2005): *Tradition through Modernity: Postmodernism and the Nation-State in Folklore Scholarship*. (Studia Fennica Folkloristica 15.) Helsinki: Finnish Literature Society.
- Anttonen, Pertti (2013): Lost in Intersemiotic Translation? The Problem of Context in Folk Narratives in the Archive. In *ARV. Nordic Year Book of Folklore*. 69. Arne Bugge Amundsen (szerk.). Uppsala, Sweden: Royal Gustavus Adolphus Academy, 153–170.
- Barna Gábor (szerk.) (2003): *Történeti források és jelenkori folkloriszövegek lejegyzésének, átírásának és kiadásának kérdései*. Szeged: SzTE Néprajzi – Kulturális Antropológiai Tanszék.
- Bárh Dániel (2012): Történeti folklorisztika Magyarországon. Előzmények és távlatok. *Etnoszkóp* (2)1: 9–21.
- Bauman, Richard (2012): Performance. In *A Companion to Folklore*. Regina Bendix és Galit Hasan-Rokem (szerk.). (Blackwell Companions to Anthropology.) Malden, MA: Wiley-Blackwell, 94–118.
- Baycroft, Timothy és David M. Hopkin (szerk.) (2012): *Folklore and Nationalism in Europe during the Long Nineteenth Century*. (National Cultivation of Culture 4.) Leiden: Brill.
- Belinko, Lital és Pavel Kats (2014): Proverbial Corpora Online. In *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet*. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5.) Meyer Holger, Christoph Schmitt, Stefanie Janssen és Alf-Christian Schering (szerk.). Münster: Waxmann, 134–142.
- Ben-Amos, Dan (1969): Analytical Categories and Ethnic Genres. *Genre* 2(3): 275–301.
- Ben-Amos, Dan (1981): Introduction. In *uő Folklore Genres*. Austin: University Of Texas Press, ix–xlv.
- Berry, David M. (2011): The Computational Turn: Thinking About The Digital Humanities. *Culture Machine* (12): 1–22.
- Berry, David M. (2012): Introduction: Understanding the Digital Humanities. In *Understanding Digital Humanities*. David M. Berry (szerk.). Basingstoke: Palgrave Macmillan, 1–20.
- Berry, David M. (szerk.) (2012): *Understanding Digital Humanities*. Basingstoke: Palgrave Macmillan.
- Beyer, Jürgen (2011): Are Folklorists Studying the Tales of the Folk? *Folklore* 122(1): 35–54.
- Bishop, Julia C. (2013): The Working Papers of Iona and Peter Opie. *Oral Tradition* 28(2): 205–216.
- Briggs, Charles L. (1988): *Competence in Performance: The Creativity of Tradition in Mexicano Verbal Art*. (University of Pennsylvania Press Conduct and Communication Series.) Philadelphia: University of Pennsylvania Press.
- Briggs, Charles L. (1993): Metadiscursive Practices and Scholarly Authority in Folkloristics. *The Journal of American Folklore* 106(422): 387–434.
- Briody, Mícheál (2007): The Schools Scheme 1937–1938; Cataloguing and archiving of material. In *uő The Irish Folklore Commission 1935–1970: History, Ideology, Methodology*. Helsinki: Finnish Literature Society, 260–270; 325–331.
- Broadwell, Peter M., David Mimno és Timothy R. Tangherlini (2014): The Tell-tale Hat: Reverse Engineering a Folklore Expert. (Előadás a *Digital Humanities '2014* konferencián, Lausanne, Switzerland.) Interneten: <http://dharchive.org/paper/DH2014/Paper-163.xml> (letöltve: 2017. január 16.).
- Dávidházi Péter (2014): Preface. Exploring Paradigms and Ourselves. In *New Publication Cultures in the Humanities: Exploring the Paradigm Shift*. Dávidházi Péter (szerk.). Amsterdam: Amsterdam University Press, 9–18.

- Debreczeni Attila (2014): Kritikai kiadás papíron és képernyőn. In *Textológia – filológia – értelmezés. Klasszikus magyar irodalom*. (Csokonai könyvtár: bibliotheca studiorum litterarium 55.) Czifra Mariann és Szilágyi Márton (szerk.). Debrecen: Debreceni Egyetemi Kiadó, 26–39.
- Dégh Linda (1986): Introduction: Special Double Issue. The Comparative Method in Folklore. *Journal of Folklore Research* 23(2–3): 77–85.
- Dixon, Dan (2012): Analysis Tool or Research Methodology: Is There an Epistemology for Patterns? In *Understanding Digital Humanities*. David M. Berry (szerk.). Basingstoke: Palgrave Macmillan, 191–209.
- Domokos Mariann (2015): A folklórgyűjtővel és folklórszövegekkel szembeni elvárások a 19. században. In *Médiák és váltások. Identitások és médiák II*. Neumer Katalin (szerk.). Budapest: Gondolat Kiadó, 30–42.
- Dundes, Alan (1997): The Motif-Index and the Tale Type Index: A Critique. *Journal of Folklore Research* 34(6): 195–202.
- Evans, Leighton és Sian Rees (2012): An Interpretation of Digital Humanities. In *Understanding Digital Humanities*. David M. Berry (szerk.). Basingstoke: Palgrave Macmillan, 21–41.
- Fine, Elizabeth C. (1984): *The Folklore Text: From Performance to Print*. Bloomington: Indiana University Press.
- Finlayson, Mark Alan (2016): Inferring Propp's Functions from Semantically Annotated Text. *The Journal of American Folklore* 129(511): 55–77.
- Finnegan, Ruth H. (1992): *Oral Traditions and the Verbal Arts: A Guide to Research Practices*. (ASA Research Methods in Social Anthropology.) London: Routledge.
- Foley, John Miles (1997 [1995]): "Folk Literature" In *Scholarly Editing: A Guide to Research*. David C. Greetham (szerk.). New York: The Modern Language Association of America, 600–626.
- Forrai Ibolya (2000): Kéziratgyűjtemény. In *A Néprajzi Múzeum gyűjteményei*. Fejős Zoltán (főszerk.). Budapest: Néprajzi Múzeum, 611–648.
- Frog (2013): Revisiting the Historical-Geographic Method(s). *RMN Newsletter* 7. Special Issue: Limited Sources, Boundless Possibilities Textual Scholarship and the Challenges of Oral and Written Texts. 18–34.
- Fumerton, Patricia és Eric Nebeker (2013): Noting the Tunes of Seventeenth-Century Broadside Ballads: The English Broadside Ballad Archive (EBBA). *Oral Tradition* 28(2): 187–192.
- Golden, Richard M. (2015): Statistical Pattern Recognition. In *International Encyclopedia of the Social & Behavioral Sciences*. 23. (Second edition). James D. Wright et al. (szerk.). Amsterdam: Elsevier, 411–417.
- Golopenția, Sanda (1997): Mapping a Network of Semiotic Systems: The Romanian Love Charms Database. *Semiotica* 114(1–2): 41–66.
- Graça da Silva, Sara és Jamshid J. Tehrani (2016): Comparative Phylogenetic Analyses Uncover the Ancient Roots of Indo-European Folktales. *Royal Society Open Science* 3(1): 150645. Interneten: <http://rsos.royalsocietypublishing.org/content/3/1/150645> (letöltve: 2017. október 4.).
- Granasztói Péter (2008): Megőrzés, hozzáférés, digitalizálás: új kihívások előtt a néprajzi archívumok. *Néprajzi Értésítő* XC. Szarvas Zsuzsa (szerk.). Budapest: Néprajzi Múzeum, 125–132.
- Granasztói Péter (2013): Az Etnológiai Archívum mint komplex gyűjtemény fejlesztésének kérdései. *Néprajzi Értésítő* XCIV. Szarvas Zsuzsa (szerk.). Budapest: Néprajzi Múzeum, 23–30.
- Gulyás Judit (2015): A szóbeliség értéke, értelmezése és a folklorisztika önmeghatározása. In *Médiák és váltások. Identitások és médiák II*. Neumer Katalin (szerk.). Budapest: Gondolat Kiadó, 11–29.
- Gunnell, Terry (2010): Sagnagrunnur: A New Database of Icelandic Folk Legends in Print. *Folklore: Electronic Journal of Folklore* (45): 151–162.
- Gunnell, Terry, Elliott Oring, Barbro Klein, John Lindow, Haim Weiss, Ülo Valk, Timothy R. Tangherlini és Fredrik Skott (2013): Discussion. Why Should Folklore Students Study "Dead" Legends? (A Round-Table Discussion Held at the 16th Congress of the International Society for Folk Narrative Research in Vilnius, Lithuania, 2013. június 29.) In *ARV. Nordic Year Book of Folklore*. 69. Arne Bugge Amundsen (szerk.). Uppsala, Sweden: Royal Gustavus Adolphus Academy, 171–209.
- Harvilahti, Lauri (2012): Finland. In *A Companion to Folklore*. (Blackwell Companions to Anthropology.) Regina Bendix és Galit Hasan-Rokem (szerk.). Malden, MA: Wiley-Blackwell, 391–408.
- Harvilahti, Lauri (2013): The SKVR Database of Ancient Poems of the Finnish People in Kalevala Meter and the Semantic Kalevala. *Oral Tradition* 28(2): 223–232.
- Herranen, Gun és Lassi Saressalo (szerk.) (1978): *A Guide to Nordic Tradition Archives*. (NIF Publications 7.) Turku: Nordic Institute of Folklore.
- Holger, Meyer, Alf-Christian Schering és Christoph Schmitt (2014): WossiDiA – The Digital Wossidlo Archive. In *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet*. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5.) Meyer Holger, Christoph Schmitt, Stefanie Janssen és Alf-Christian Schering (szerk.). Münster: Waxmann, 61–85.

- Holger, Meyer, Christoph Schmitt, Stefanie Janssen és Alf-Christian Schering (szerk.) (2014): *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet*. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5.) Münster: Waxmann.
- Honko, Lauri (1998): *Textualising the Siri Epic*. (FF Communications 264.) Helsinki: Suomalainen Tiedekatemia Academia Scientiarum Fennica.
- Honko, Lauri (2000a): Text as Process and Practice: The Textualization of Oral Epics. In *The Textualization of Oral Epics*. Lauri Honko (szerk.). The Hague: Mouton, 3–56.
- Honko, Lauri (2000b): Thick Corpus, Organic Variation: an Introduction. In *Thick Corpus, Organic Variation and Textuality in Oral Tradition*. (Studia Fennica Folkloristica 7.) Lauri Honko (szerk.). Helsinki: Finnish Literature Society, 3–28.
- Honko, Lauri (2001): The digital era is here. *FF Network* 22: 1, 13.
- Hughes, Lorna, Panos Constantopoulos és Costis Dalla (2016): Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 150–170.
- Ilyefalvi Emese (2017): Textualizációs eljárások, tipológiai kísérletek, digitális adatbázisok. Az összehasonlító szövegfolklorisztikai kutatások jövője a ráolvasások tükrében. *Ethnographia* 129(3): 383–415.
- Karsdorp, Folgert, Marten van der Meulen, Theo Meder és Antal van den Bosch (2015): MOMFER: A Search Engine of Thompson's Motif-Index of Folk Literature. *Folklore* 126(1): 37–52.
- Katajamäki, Sakari és Karina Lukin (2013): Textual Trails from Oral to Written Sources: An Introduction. *RMN Newsletter* 7. Special issue: Limited Sources, Boundless Possibilities Textual Scholarship and the Challenges of Oral and Written Texts. 8–17.
- Kenna, Ralph, Máirín MacCarron és Pádraig MacCarron (szerk.) (2017): *Maths Meets Myths: Quantitative Approaches to Ancient Narratives*. (Understanding Complex Systems.) Springer International Publishing.
- Keszeg Vilmos (2011): *Alfabetizáció, írásszokások, populáris írásbeliség* (Néprajzi Egyetemi Jegyzetek 3.) Kolozsvár: Kriza János Néprajzi Társaság – BBTE Magyar Néprajz és Antropológia Tanszék.
- Kikas, Katre (2014): Folklore Collecting as Vernacular Literacy: Establishing a Social Position for Writing in the 1890s Estonia. In *Vernacular literacies. Past, Present and Future*. Ann-Catrine Edlund, Susanne Haugen és Lars-Erik Edlund (szerk.). Umeå: Umeå University, 221–235.
- Kokas Károly (2016): Digitális bölcsészet 2016. A bölcsészek és az informatikai megközelítés: régen és most. In *MONOKgraphia. Tanulmányok Monok István 60. születésnapjára*. Nyerges Judit, Verók Attila és Zvara Edina (szerk.). Budapest: Kossuth Kiadó, 405–412.
- Kolovos, Andy (2004): Contextualizing the Archives. *Folklore Forum* 35(112): 18–28.
- Kolovos, Andy (2010): *Archiving Culture: American Folklore Archives in Theory and Practice*. (Doktori disszertáció) Department of Folklore and Ethnomusicology, Indiana University: Bloomington.
- Kõiva, Mare (2003) Folkloristics Online. The Estonian Experience. *Folklore: Electronic Journal of Folklore* 25: 7–34.
- Kulasalu, Kaisa (2013): Immoral Obscenity: Censorship of Folklore Manuscript Collections in Late Stalinist Estonia. *Journal of Ethnology and Folkloristics* 7(1): 66–81.
- Kuutma, Kristin (2015): From Folklore to Intangible Heritage. In *A Companion to Heritage Studies*. (Blackwell Companions to Anthropology) William Logan, Máiréad Nic Craith és Ullrich Kockel (szerk.). Hoboken, NJ: John Wiley & Sons, 41–54.
- Järv, Risto (2013): Estonian Folklore Archives. *Oral Tradition* 28(2): 291–298.
- Järv, Risto és Mari Sarv (2014): From Regular Archives to Digital Archives. In *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet*. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5.) Meyer Holger, Christoph Schmitt, Stefanie Janssen és Alf-Christian Schering (szerk.). Münster: Waxmann, 49–60.
- Jockers, Matthew L. (2013): *Macroanalysis: Digital Methods and Literary History* (Topics in the Digital Humanities.) Urbana – Chicago – Springfield: University of Illinois Press.
- Jockers, Matthew L. és Ted Underwood (2016): Text Mining and the Humanities. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 291–306.
- Labádi Gergely (2014): A filológiai tudás formái. In *Textológia – filológia – értelmezés. Klasszikus magyar irodalom*. (Csokonai könyvtár: bibliotheca studiorum litterarium 55.) Czifra Mariann és Szilágyi Márton (szerk.). Debrecen: Debreceni Egyetemi Kiadó, 173–190.



- Lajoie, Patrice, Julien d'Huy és Jean-Loïc Le Quellec (2013): Comments on Tehrani. *Nouvelle Mythologie Comparée/ New Comparative Mythology* Interneten: <http://nouvellemythologiecomparee.hautetfort.com/archive/2013/12/04/patrice-lajoie-julien-d-huy-and-jean-loic-le-quellec-comment-5237721.html> (letöltve: 2017. január 16.).
- Landgraf Ildikó (2006): Archívumon innen, katalóguson túl. Többletek és hiányok a mai magyar történeti mondatkutatás műfajelméleti és rendszerezési kérdéseiben. In *Ethno-lore*. XXIII. Hoppál Mihály, Vargyas Gábor és Berta Péter (szerk.). Budapest: Akadémiai, 27–40.
- Landgraf Ildikó (2016): A 24. óra szorításában. A folklorgyűjtés alapelvei Katona Lajos munkásságában és a korabeli magyar folklorisztikában. *Ethnographia* 127(4): 503–519.
- Lin, Yu-Wei (2012): Transdisciplinarity and Digital Humanities: Lessons Learned from Developing Text-Mining Tools for Textual Analysis. In *Understanding Digital Humanities*. David M. Berry (szerk.). Basingstoke: Palgrave Macmillan, 295–314.
- Mahlamäki, Tiina (2001): Some guidelines for the archiving of qualitative research data in the digital era. *FF Network* 22: 2–5, 13.
- Manovich, Lev (2009 [2001]): Az adatbázis mint szimbolikus forma. *Apertúra* 5(1) Interneten: <http://uj.apertura.hu/2009/osz/manovich/> (letöltve: 2017. augusztus 23.).
- Markoff, John (2015): Archival Methods. In *International Encyclopedia of the Social & Behavioral Sciences*. 1. (Second edition). James D Wright et al. (szerk.). Amsterdam: Elsevier, 909–915.
- McCarty, Willard (2016): Becoming Interdisciplinary. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 69–83.
- McGann, Jerome J. (2010): Electronic Archives and Critical Editing. *Literature Compass* 7(2): 37–42.
- McGann, Jerome J. (2014): *A New Republic of Letters: Memory and Scholarship in the Age of Digital Reproduction*. Cambridge, Massachusetts: Harvard University Press.
- McGann, Jerome J. (2016): Marking Texts of Many Dimensions. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 358–376.
- Meder, Theo (2010): From a Dutch Folktale Database towards an International Folktale Database. *Fabula* 51(1–2): 6–22.
- Meder, Theo (2014a): The Folktale Database as a Digital Heritage Archive and as a Research Instrument. In *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet*. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5.) Meyer Holger, Christoph Schmitt, Stefanie Janssen és Alf-Christian Schering (szerk.). Münster: Waxmann, 119–129.
- Meder, Theo (2014b): *Committee for «Folktales and the Internet»*. Interneten: <http://www.isfnr.org/files/CommitteeInternet.pdf> (letöltve: 2017. január 18.).
- Meder, Theo, Folgert Karsdorp, Dong Nguyen, Mariët Theune, Dolf Trieschnigg és Iwe Everhardus Christiaan Muiser (2016): Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database. *Journal of American Folklore* 129(511): 78–96.
- Mikkola, Kati (2013): Self-Taught Collectors of Folklore and Their Challenge to Archival Authority. In *White Field, Black Seeds. Nordic Literacy Practices in the Ling Nineteenth Century*. (Studia Fennica Litteraria 7.) Anna Kuismin és M. J. Discroll (szerk.). Helsinki: Finnish Literature Society, 146–157.
- Moretti, Franco (2000): Conjectures on World Literature. *New Left Review* 1 Interneten: <https://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature> (letöltve: 2017. szeptember 28.).
- Moretti, Franco (2007): *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.
- Muiser, Iwe Everhardus Christiaan, Mariët Theune és Theo Meder (2012): Cleaning up and Standardizing a Folktale Corpus for Humanities Research. In *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH2)* Francesco Mambrini, Marco Passarotti és Caroline Sporleder (szerk.). Lisboa: Edições Colibri, 63–74.
- Munk, Anders Kristian és Torben Elgaard Jensen (2014): Revisiting the Histories of Mapping. Is there a Future for a Cartographic Ethnology? *Ethnologia Europaea* Special issue: European Ethnology Revisited. 44(2): 31–47.
- Nic Craith, Máiréad (2008): From National to Transnational. A Discipline en route to Europe. In *Everyday Culture in Europe: Approaches and Methodologies. Progress in European Ethnology*. Máiréad Nic Craith, Ullrich Kockel és Reinhard Johler (szerk.). Aldershot: Ashgate, 1–17.
- NIF (Nordic Institute of Folklore) Newsletter* 1978 6(1); 1982 10(4); 1989 17(4); 1995 (23)4.
- Niles, John D. (2013a): Orality. In *The Cambridge Companion to Textual Scholarship*. (Cambridge Companions to Literature.) Neil Fraistat és Julia Flanders (szerk.). Cambridge: Cambridge University Press, 215–223.
- Niles, John D. (2013b): From Word to Print – and Beyond. *Western Folklore* 72(3–4): 229–251.



- Ógáin, Ríonach (2013): Cnuasach Bhéaloideas Éireann: The National Folklore Collection, University College Dublin. *Oral Tradition* 28(2): 317–324.
- Paßmann, Johannes és Asher Boersma (2017): Unknowing Algorithms on Transparency of Unopenable Black Boxes. In *The Datafied Society. Studying Culture through Data*. Mirko Tobias Schäfer és Karin van Es (szerk.). Amsterdam: Amsterdam University Press, 139–146.
- Pierazzo, Elena (2016): Textual Scholarship and Text Encoding. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 307–321.
- Presner, Todd és David Shepard (2016): Mapping the Geospatial Turn. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 201–212.
- Presner, Todd, David Shepard és Yoh Kawano (2014): *Hypercities Thick Mapping in the Digital Humanities*. Cambridge, Mass: Harvard University Press.
- Price, Kenneth M. (2016): Social Scholarly Editing. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 137–149.
- Rajamäki, Maria (1989): Introducing Colcard. *NIF Newsletter* 17(4): 35–39.
- Rieder, Bernhard és Theo Röhle (2012): Digital Methods: Five Challenges. In *Understanding Digital Humanities*. David M. Berry (szerk.). Basingstoke: Palgrave Macmillan, 67–84.
- Riesch, Hauke (2015): Citizen Science. In *International Encyclopedia of the Social & Behavioral Sciences*. (Második kiadás.) 3. kötet. James D. Wright (főszerk.). Amsterdam: Elsevier, 631–636.
- Rogan, Bjarne (2012): The Institutionalization of Folklore. In *A Companion to Folklore*. (Blackwell Companions to Anthropology.) Regina Bendix és Galit Hasan-Rokem (szerk.). Malden, MA: Wiley-Blackwell, 598–630.
- Rogan, Bjarne (2014): Popular Culture and International Cooperation in the 1930s. CIAP and the League of Nations. In *Networking the International System. Transcultural Research – Heidelberg Studies on Asia and Europe in a Global Context*. Madeleine Herren (szerk.). Springer International Publishing, 175–185.
- Rosenstock, Bruce és Belén Bistué (2013): The Folk Literature of the Sephardic Jews Digital Library. *Oral Tradition* 28(2): 325–334.
- Ryan, Catherine (2014a): *Thesaurus Construction Guidelines: An Introduction to Thesauri and Guidelines on Their Construction*. Dublin: Royal Irish Academy and National Library of Ireland.
- Ryan, Catherine (2014b): *Report on the MoTIF Project: Thesaurus Guidelines and Pilot Thesaurus of Irish Folklore*. Dublin: Royal Irish Academy and National Library of Ireland.
- Saarinen, Jukka (2001): Kalevalaic Poetry as a Digital Corpus. *FF Network* (22): 6–9.
- Seifert, Manfred és Hendrik Keller (2014): Adolf Spamer online Vorüberlegungen zu einem Projekt der volkskundlich-kulturwissenschaftlichen Bestandssicherung und Öffentlichkeitsarbeit. In *Corpora ethnographica online: Strategien der Digitalisierung Kultureller Archive und ihrer Präsentation im Internet*. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5.) Meyer Holger, Christoph Schmitt, Stefanie Janssen és Alf-Christian Schering (szerk.). Münster: Waxmann, 85–100.
- Seitel, Peter (2012): Three Aspects of Oral Textuality. In *A Companion to Folklore*. (Blackwell Companions to Anthropology.) Regina Bendix és Galit Hasan-Rokem (szerk.). Malden, MA: Wiley-Blackwell, 75–93.
- Schäfer, Mirko Tobias és Karin van Es (2017): Introduction. New Brave World. In *The Datafied Society. Studying Culture through Data*. Mirko Tobias Schäfer és Karin van Es (szerk.). Amsterdam: Amsterdam University Press, 13–22.
- Schäfer, Mirko Tobias és Karin van Es (szerk.) (2017): *The Datafied Society. Studying Culture through Data*. Amsterdam: Amsterdam University Press.
- Schreibman, Susan, Raymond George Siemens és John Unsworth (szerk.) (2016): *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Chichester, West Sussex: Wiley Blackwell.
- Shillingsburg, Peter (2016): Reliable Social Scholarly Editing. *Digital Scholarship in the Humanities* 31(4): 890–897.
- Shuman, Amy és Galit Hasan-Rokem (2012): The Poetics of Folklore. In *A Companion to Folklore*. (Blackwell Companions to Anthropology.) Regina Bendix és Galit Hasan-Rokem (szerk.). Malden, MA: Wiley-Blackwell, 55–74.
- Sinclair, Stéfan és Geoffrey Rockwell (2016): Text Analysis and Visualization: Making Meaning Count. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 274–290.
- Skott, Fredrik (2001): „Most of your questionnaires are terrible to work with.” In *Input & Output: The Process of Fieldwork, Archiving and Research in Folklore*. (NNF Publications 10.) Ulrika Wolf-Knuts, Anders Salomonsson, Ann Helene Bolstad Skjelbred és Ríonach uí Ógáin (szerk.). Turku: Nordic Network of Folklore, 75–114.

- Skott, Fredrik (2008): Summary. In uő *Folkets minnen: traditionsinsamling i idé och praktik 1919–1964*. (Avhandlingar från Historiska institutionen i Göteborg 53.) Göteborg: Institutet för språk och folkminnen i samarbete med Göteborgs universitet, 281–285.
- Szilágyi Márton (2014): Textológia, filológia, értelmezés. In *Textológia – filológia – értelmezés. Klasszikus magyar irodalom*. (Csokonai könyvtár: bibliotheca studiorum litterarium 55.) Czifra Mariann és Szilágyi Márton (szerk.). Debrecen: Debreceni Egyetemi Kiadó, 15–25.
- Tangherlini, Timothy R. (2013a): The Folklore Macroscope. *Western Folklore* 72(1): 7–27.
- Tangherlini, Timothy R. (2013b): *Danish Folktales, Legends, and Other Stories*. (New Directions in Scandinavian Studies.) Seattle – Copenhagen: University of Washington Press – Museum Tusulanum Press.
- Tangherlini, Timothy R. (2014): Sites of (re)Collection: Creating the Danish Folklore Nexus. *Journal of Folklore Research* 51(2): 223–247.
- Tangherlini, Timothy R. (2016a): Big Folklore: A Special Issue on Computational Folkloristics. *Journal of American Folklore* 129(511): 5–13.
- Tangherlini, Timothy R. (2016b): WitchHunter: Tools for the Geo-Semantic Exploration of a Danish Folklore Corpus. *Journal of American Folklore* 129(511): 14–42.
- Tangherlini, Timothy R. és Peter M. Broadwell (2017): GhostScope: Conceptual Mapping of Supernatural Phenomena in a Large Folklore Corpus. In *Maths Meets Myths: Quantitative Approaches to Ancient Narratives*. (Understanding Complex Systems.) Ralph Kenna, Máirín MacCarron és Pádraig MacCarron (szerk.). Springer International Publishing, 131–157.
- Tangherlini, Timothy R. és Peter Leonard (2013): Trawling in the Sea of the Great Unread: Sub-corpus Topic Modeling and Humanities Research. *Poetics* 41(6): 725–749.
- Tehrani, Jamshid J. (2013a): The Phylogeny of Little Red Riding Hood. *PLoS ONE* 8(11):e79971: 1–11. Interneten: <http://dx.doi.org/10.1371/journal.pone.0078871> (letöltve: 2017. január 16.).
- Tehrani, Jamshid J. (2013b): Reply to Lajoie, d’Huy and Le Quellec. *Nouvelle Mythologie Comparée/New Comparative Mythology* Interneten: <http://nouvellemythologiecomparee.hautetfort.com/archive/2013/12/11/jamshid-j-tehrani-reply-to-lajoie-d-huy-and-le-quellec-2013-5244250.html> (letöltve: 2017. január 16.).
- Tehrani, Jamshid J., Quan Nguyen és Teemu Roos (2015): Oral Fairy Tale or Literary Fake? Investigating the Origins of Little Red Riding Hood Using Phylogenetic Network Analysis. *Digital Scholarship in the Humanities* 31(3): 611–636. Interneten: <http://dx.doi.org/10.1093/lc/fqv016> (letöltve: 2017. január 16.).
- Tehrani, Jamshid J. és Julien d’Huy (2017): Phylogenetics Meets Folklore: Bioinformatics Approaches to the Study of International Folktales. In *Maths Meets Myths: Quantitative Approaches to Ancient Narratives*. (Understanding Complex Systems.) Ralph Kenna, Máirín MacCarron és Pádraig MacCarron (szerk.). Springer International Publishing, 91–114.
- Terras, Melissa (2016): Crowdsourcing in the Digital Humanities. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 420–438.
- Thomas, William G. III. (2016): The Promise of the Digital Humanities and the Contested Nature of Digital Scholarship. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 524–537.
- Valk, Ülo (2005): Establishment of the Estonian Folklore Collections and the Concept of Authenticity. In *Völkundliche Großprojekte: Ihre Geschichte und Zukunft*. Christoph Schmitt (szerk.). Münster: Waxmann, 33–38.
- Vargha Katalin (2016): A digitális folklorisztika felé. Egy új kulcsszó és háttere a nemzetközi kutatásban. *Ethnographia* 127(4): 624–637.
- Västrik, Ergo-Hart (2007): *Archiving Tradition in a Changing Political Order: From Nationalism to Pan-Finno Ugrianism in the Estonian Folklore Archives* (Paper prepared for the conference “Culture Archives and the State: Between Nationalism, Socialism, and the Global Market,” May 3–5, 2007, Mershon Center, Ohio State University, USA). Interneten: [https://kb.osu.edu/dspace/bitstream/handle/1811/46903/1/FolkloreCntr\\_2007conference\\_Vastrik7.pdf](https://kb.osu.edu/dspace/bitstream/handle/1811/46903/1/FolkloreCntr_2007conference_Vastrik7.pdf) (letöltve: 2017. január 16.).
- Virtanen, Lea (1993): Is There the Comparative Method out of Date? In *Telling Reality. Folklore Studies in Memory of Bengt Holbek*. Michael Chesnutt (szerk.). Copenhagen és Turku: Department of Folklore, University of Copenhagen, 255–272.
- Voigt Vilmos és Balogh Lajos (1974): *A népköltési (folklór) alkotások kritikai kiadásának szabályzata*. Budapest: Akadémiai.
- Voigt Vilmos (1981): Computertechnik und -analyse. In *Enzyklopädie des Märchens Handwörterbuch zur historischen und vergleichenden Erzählforschung*. 3. Kurt Ranke (főszerk.). Berlin és New York: Walter de Gruyter, 111–123.

- Voigt Vilmos (1997): Megoldott és megoldatlan kérdések hangrögzítésünk kezdetei körül. A kép- és hangrögzítés változó módszerei a néprajzi kutatásban. *Néprajzi Értesítő* LXXIX. Selmeczi Kovács Attila (szerk.). Budapest: Néprajzi Múzeum, 103–107.
- Voigt Vilmos (2004): A magyar folklór textológia helyzete és új távlatai. *Irodalomtörténet* 85(3): 356–366.
- Voigt Vilmos (2006): The Theory of Database in Folk Narrative Studies. *Fabula* 47(3–4): 308–318.
- Warwick, Claire (2016): Building Theories or Theories of Building? A Tension at the Heart of Digital Humanities. In *A New Companion to Digital Humanities*. (Revised edition.) (Blackwell Companions to Literature and Culture.) Susan Schreibman, Raymond George Siemens és John Unsworth (szerk.). Chichester, West Sussex: Wiley Blackwell, 538–552.
- Wolf-Knuts, Ulrika (2000): On the History of Comparison in Folklore Studies In *Thick Corpus, Organic Variation and Textuality in Oral Tradition*. (Studia Fennica Folkloristica 7.) Lauri Honko (szerk.). Helsinki: Finnish Literature Society, 254–283.
- Wolf-Knuts, Ulrika (2001): Cultural Conditions for Fieldwork and Archiving. In *Input & Output: The Process of Fieldwork, Archiving and Research in Folklore*. (NNF Publications 10.) Ulrika Wolf-Knuts et al. (szerk.). Turku: Nordic Network of Folklore, 9–24.
- Wynne, Martin (2012): Do We Need Annotated Corpora in the Era of the Data Deluge? (Keynote abstract.) In *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities*. (ACRH2) Francesco Mambrini, Marco Passarotti és Caroline Sporleder (szerk.). Lisboa: Edições Colibri, 1–2.