Acta Ethnographica Hungarica 63(1), 209–258 (2018)
DOI: 10.1556/022.2018.63.1.11

The Theoretical, Methodological and Technical Issues of Digital Folklore Databases and Computational Folkloristics[1]

Emese Ilyefalvi
Junior Research Fellow: "East–West" ERC Research Group, Institute of Ethnology, RCH, Hungarian Academy of Sciences, Budapest;
PhD Candidate: Eötvös Loránd University, Department of Folklore

Abstract: The study examines the problems and possibilities presented by the digitization of national folklore archives and collections in the wider context of folklore archiving and digital humanities. The primary goal of the study is to present a problem-oriented and critical overview of the available digital databases containing folklore texts (WossiDiA, Sagragrunnur, ETKSpace, Danish Folklore Nexus, Nederlandse VolksverhalenBank, The Schools' Collection, etc.), and of the analyses conducted on these using computational methods. The paper first presents a historical overview of the conceptualization that went into the creation of folklore databases (genre-centered, collector, and collection-centered approaches), followed by a discussion of the practical, technical, and theoretical aspects of digital content creation (crowdsourcing, markup languages, TEI, digital critical editions, etc.). The study then takes a look at the new digital tools and methods applied in the analysis of digitized folklore texts (text-mining, network theory methods, data visualization), and finally places databases and computational folkloristics within a larger theoretical framework.
Keywords: computational folkloristics, digital databases, digital editing, folklore archives, folklore texts, folklore genres, history of European folklore archives, textualization, theory and methods of digitization

[1] The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement № 324214, and Supported By Únkp-17-3 New National Excellence Program Of The Ministry Of Human Capacities.

DIGITAL HUMANITIES, DIGITAL TEXTOLOGY AND FOLKLORISTICS[2]

Over the past twenty years, digital technology has almost inconspicuously reshaped our everyday lives which, according to historians of science and philosophers, is a more significant turning point in human culture than the invention of printing during the Renaissance (Dávidházi 2014; McGann 2014). This medium has radically altered the limits and possibilities of scientific research. Although the possibilities for practical application are self-evident – that is, today it is practically unimaginable to conduct research that does not in some manner make use of computers or of other opportunities offered by the digital world – it is still debatable whether it will result in a new methodology, a paradigm shift, or perhaps a new, independent discipline within the humanities.[3] The processes taking place before our eyes occur in the way of previous scientific revolutions, characterized by competing methods, theories and terminological turmoil,[4] the end result for now uncertain and unknown.[5] At the same time, compared to the past, all of this is happening at the speed of light. While digital technology has fundamentally displaced most human science disciplines from their well-established positions (which may be why many in the humanities look at it with suspicion and reluctance), the newly emerged discipline, referred to as *Digital Humanities* in the Anglo-Saxon language area, has built up its own institutional system (through international societies, annual congresses, etc.),[6] and its international historiographers are already discussing a third wave within the discipline.[7]

At this point, the digital revolution(s) has primarily affected the philological-textological work in the humanities. This includes the basic activities of text production and text edition, which becomes evident through a quick survey of publications and conferences related to the digital humanities, a majority of which deal with the issues of digital textology (McGann 2014; Dávidházi 2014; Debreczeni 2014; Schreibman et

[2] The lecture outlines of the current study have previously been presented at the Folklore Fellows' 2015 Summer School called "Doing Folkloristics in the Digital Age" in Turku, at the 2015 Congress of the SIEF Working Group on Archives in Zagreb, and at the international conference of the Kriza János Ethnographic Society ("The Changing Contexts of Archive Use") on October 14–15, 2016 in Otomani, Romania. I wish to express my gratitude for the many contributions and constructive criticisms I received at the events mentioned above. I am also grateful to all the members of the "East–West" Research Group at the Institute of Ethnology of the Research Centre for the Humanities of the Hungarian Academy of Sciences, who have commented on the earlier versions of the text.

[3] Kokas 2016: 405; Thomas 2016. Recent handbooks and overviews of digital humanities: Schreibman – Siemens – Unsworth (eds.) 2016; Berry (ed.) 2012.

[4] When it comes to folkloristics, the two terms used are *digital* and *computational folkloristics*. Cf. Vargha 2016.

[5] From the perspective of the history of science, Claire Warwick compared the institutionalization of digital humanities to the 19th-century struggle of English Studies Warwick 2016.

[6] For the institutional system of digital humanities, see Thomas 2016.

[7] The first wave of the late 1990s and early 2000s was characterized by a qualitative approach, i.e., the goal was to digitize more material and develop an appropriate infrastructures for analyzing these large text corpora. Researchers considered digital technology as a supplementary, accelerator tool. With regard to the use of this term, this period was characterized by *computing in the humanities* and *humanities computing*. In the second wave, however, with the name change – *digital humanities* – came autonomous disciplinary needs, namely the use of specific and hybrid methods and tools, publishing models that are no longer based on book culture, and the focus was on the problem of inherently digital materials. Berry 2011:2–4, 2012.

al. (eds.) 2016). The digital revolution has already benefited the way in which textology's position has been perceived as merely an auxiliary, "handmaiden" science of the humanities (McGann 2014:19–20; Szilágyi 2014).[8] As the static, analog critical editions could not fit the various postmodern textual theories, the digital medium, with its dynamic and flexible possibilities, is proving to be a more suitable tool (Debreczeni 2014:27–28).[9]

The aforementioned transformations also compell folkloristics to refocus on essential issues and fundamental questions of the discipline. Methodological and theoretical dilemmas stemming from data collection, archiving, textological procedures and techniques have been present since the beginning of folklore studies (Cf. Gulyás 2015; Landgraf 2016) resulting in a lively discourse that has in the course of the 150-year history of the discipline led to several paradigm shifts.[10] That is why, in the evolving paradigm of digital humanities, it is necessary to debate the following: 1. How should the vast amount of textual material accumulated by folklore research be stored and published? 2. What tools and procedures are needed for this? 3. What new methods do we need for analysis? 4. What kinds of analyses are digital folklore text corpora suitable or not suitable for?[11]

In my study, I survey the issues and opportunities of the digitization of certain national folklore archives and collections that were created during the institutionalization of European ethnography and folkloristics in the late 19th and early 20th centuries.[12] The primary purpose is to provide an issue-oriented and critical overview of the available digital databases containing and processing folklore texts and their analyses.[13] I discuss the question in the broader context of folklore archiving and digital humanities. First I present a historical overview of the conceptualization going into the creation of folklore databases, and then the practical, technical and methodological-theoretical aspects of digital content creation. Following this, I examine the current digital tools and techniques used to analyze digitized folklore texts, and finally I summarize the theoretical principles of databases and computational folkloristics. Prior to that, a brief look at European folklore archives and the relationship between folkloristics and archiving is given, since

[8] The tension between theoretical and practical approaches is also found in the digital humanities: Warwick 2016.

[9] The great critics of digital textology and humanities (such as Stanley Fish) disagree with this, objecting to the tighter text limits that computer analysis necessitates and the inflexibility of the sectioning of texts and the mechanization of the concept of the text. Cf. McGann 2014, 2016.

[10] For the textualization paradigms of folkloristics, see Fine 1984; Foley 1997; Voigt 2004; Honko 2000a; Landgraf 2006; Seitel 2012; Niles 2013a, 2013b; Katajamäki – Lukin 2013. For a guide to Hungarian folklore textology: Voigt – Balogh 1974; Barna (ed.) 2003.

[11] For the tasks of computational folkloristics, see Timothy R. Tangherlini's programmatic study: Tangherlini 2013.

[12] In the present text, I refer to the archives and repositories containing folklore materials as *folklore archives*, the denomination of which varies by country (ethnographic, folkloric, ethnological, traditional). With this term I also emphasize that my approach in this study is essentially folkloristic, it comments on the digitization of manuscripts and typewritten texts of formerly oral texts.

[13] Therefore, I will not cover the theoretical, technical and ethical issues of the storage of other file formats (audio, video, photographs, etc.) and the issue of the preservation and patrimonization of minor, endangered cultures, where data gathering takes place roughly simultaneously with its archiving. Moreover, I will only include scientific databases that are machine-readable and follow interpretable and analyzable textualization procedures and are thus suitable tools for scientific research.

most digital folklore databases undertake the joint digitization of a previously created folklore archive, or part of an archive's material, or the material of several archives.

ARCHIVES AND ARCHIVING IN FOLKLORISTICS

Among the various historical repositories, manuscript libraries and archives, European folklore archives are a distinctive phenomenon.[14] I would like to refer briefly to just a few important factors to illustrate their heterogeneity.[15] The first is the diverse motivation and habitus of collecting folklore, as well as the variety of ideologies behind it (enlightenment, public education, nation building, nationalism, preservation, heritagization, traditionality, scientific research) (Anttonen 2005; Wolf-Knuts 2001, 2010; Valk 2005; Baycroft – Hopkin 2012; Kuutma 2015). Even if we were to unbundle from these various motivations the folklore collections created by scientific interest in the narrower sense, which themselves are constantly changing,[16] we still get a wide variety and diversity of material due to the plurality of folklore definitions and folklore collecting techniques that

[14] Various modern-day state, governmental, historical, and institutional public archives dating back to the French Revolution also differ widely in international comparison: Markoff 2015.

[15] There is no overview of the history of European folklore archives so far, so I only refer to a few of the issues that are most relevant to digitization. These cannot be generalized for all folklore archives, as the problem does not appear to the same extent, weight, and in the same time period in specific national cases. Regina Bendix and Galit Hasan-Rokem's handbook, *A Companion to Folklore*, somewhat compensates for this deficit with Bjarne Rogan's chapter on the institutions of European folkloristics (mentioning Northern, Irish, Dutch, French, German archives) (Rogan 2012:610–614). Andy Kolovos' dissertation on American folklore archives briefly refers to their European background and the main differences between European and American folklore archives. Cf. Kolovos 2010:1–87. An important catalog-like overview of Northern European folklore archives (Denmark, Faroe Islands, Sweden, Norway, Iceland, Finland) was published in 1978: Herranen – Saressalo (eds.) 1978.

[16] Although in terms of the material of folklore archives, the separation of *amateur* and *professional* collectors is in fact in most cases not very redeeming or even possible, from the 19th century onward, the vast majority of their material is a result of the enthusiastic work of avid volunteers or paid amateur collectors. According to Ibolya Forrai (1998), the Ethnological Archives of the Museum of Ethnography in Budapest for example, is made up of mostly material that has been submitted through the volunteer collecting network (more than 12,000 items). Cf. Forrai 2000:33. Networks of volunteer collectors, however, were not necessarily always guided by a stringent top management and a uniform collection concept, or if so, they were very different in practice when they were actually implemented. For this reason, professional practitioners were rightly afraid of dilettantization and tried to guide collections in the direction of professionalism. Lajos Katona, who played a significant role in Hungarian folkloristics becoming a science, drew attention to this at the very end of the 19th century. Cf. Landgraf 2016:509–510. For the control of Irish full-time and part-time collectors, cf. Briody 2007:415–429. Or for the same, cf. Kati Mikkola's study of the conflicts between volunteer collectors of the Finnish folklore archive and professional folklorists. The latter, however, provides an opportunity to explore the motivations and folklore concepts of self-taught volunteer collectors: Mikkola 2013. For the 19th-century interpretation of the activities of amateur, volunteer Estonian folklore collectors as vernacular literacy, cf. Kikas 2014.

exist side by side.[17] In the case of scientifically verified institutions with a long history, the transforming and generally expanding, definitions of folklore and the various methods used to preserve folkloric material have significantly transformed the nature of material identified as folklore and intended for folklore archives.[18] Folklore archives are therefore special phenomena in that many of them are a by-product of a specific research project, and hence their main influencing factor is how the research person or group determined the purpose and objective of the research and what method(s) were chosen to achieve this (Wolf-Knuts 2001:9–14). It then follows that the types of source documents are not homogenous. The collections of collectors or research programs are made up of a variety of material, such as memos, questionnaires, notes, memoirs, field logs, other ancillary elements of fieldwork, private or official correspondence, and various drafts. It is also worth pointing out that European folklore archives operate in a rather varied institutional background and system, and this has had a significant influence on the archiving structure of the collections, and, more relevantly, also on the options and frameworks of their digitization.[19] Folklore archives may be a part of university departments, or they may belong to different cultural, literary, ethnographic or folklore societies, organizations, or manuscript libraries, while elsewhere they may operate in the framework of a museum (literary, cultural or explicitly ethnographic), but they may also be a completely independent institution (Rogan 2012:610). Finally, it is important to mention the current political situation of the institutions maintaining the archives, that is, the relation of the collection to the current national government, as this has also greatly shaped and transformed their fate. The notion of a completely independent science is an illusion.[20]

Despite the differences outlined above, European folklore archives are nonetheless linked to each other in many ways, particularly because of the comparative perspectives

[17] There are numerous examples of coexisting plurality. During the transformation of the paradigm of textology in the late 19th, early 20th century, Hungarian folklorists, for example, expressed their many different opinions on the various advantages and disadvantages of collecting by memory, dictation, shorthand, or phonograph transcription, each considering one or the other as more reliable (Gulyás 2015; Landgraf 2016:511–513). Mariann Domokos points out that even the concept of collector has changed in turn-of-the-century Hungarian folkloristics, when collectors became mere documenters instead of authors (Domokos 2015). For the same, see Fredrik Skott's research on the Swedish history of the discipline. Skott pointed out that in 1930s Sweden, during the preliminary work on the Swedish ethnographic atlas, the heated debate between the Västsvsenska Folkminnesarkivet in Gothenburg (then lead by Carl-Martin Bergstrand) and the folklore archives in Uppsala, Lund, and Stockholm resulted in the emergence of two completely opposite collecting and archiving methods and concepts (Skott 2001, 2008).

[18] Since the 1960s, the Finnish folklore archive has focused on collecting life stories through their oral history campaigns, mainly from different occupations (e.g., hospital workers, road builders, etc.): Harvilahti 2012:402–404. In the context of Hungarians in Romania, cf. the collections of life histories of peasants, and later of teachers and engineers, starting from the second half of the 20th century: Keszeg 2011:165–194.

[19] From the perspective of sustainability, it is very important that databases also be part of an institutional structure (Debreczeni 2014:29), which requires continuous financial resources (Järv – Sarv 2014:59). At the same time, the nature and potential of database building are determined by the digitization objectives and priorities of the given institution.

[20] For example, the establishment of the Irish Folklore Commission's archives at the time of the birth of independent Ireland (Briody 2007:33–39). Fortunately, more researchers are studying the relationship between the Soviet era and folklore archives (Västrik 2007; Kulasalu 2013).

of folkloristics. Ethnographic collections have imagined themselves from the beginning as requiring international cooperation. The national archives are an important base of these comparative endeavors, complemented by the possiblity of a European folklore archive; this, however, never materialized (Nic Craith 2008; Rogan 2012:604–606, 2014:174). The largest intersection of international folkloristics have been the historic-geographic trends and methods that have long defined the discipline, with which large joint European ethnographical projects can be associated, such as the *Enziklopädie des Märchens*, international catalogs, or national and international ethnographic and dialectological and linguistic atlases (Schmitt (ed.) 2005; Rogan 2014:176–177). Due to the nature of volunteer collecting networks which were developed specifically for accomplishing such repositories, and the collection questionnaires, guides, and catalog systems developed after the North European model, the character and structure of the core collections of European archives are highly similar, which, despite the differences, allows the collective discussion of the theoretical, methodological and technical issues of their digitization.[21]

Archives were key to the institutionalizaton of folkloristics and played a vital role in the discipline until the mid-20th century (Gulyás 2015:18). However, the epistemological revolutions of the 1960s and 1970s significantly altered the role and value of folklore archives, especially in northern and western European and American folkloristics, which resulted in generations of scholars refusing to use such collections (Wolf-Knuts 2001:12; Kolovos 2010:23; Gunnell 2013:173; Rogan 2012:613–614; Anttonen 2013; Harvilahti 2012:402–403).[22] As the analog book did not fit postmodern textual theories, so the anthropological and pragmatic revolution in folkloristics resulted in the archive and its rigid structure not meeting the transformed needs of the discipline.[23] The attention in folkloristics shifted from the past to the present, from text to performance and use, from structure and format to context and interaction, from the community to the individual, from rural to urban, from oral to written, and consequently from the *archive* to the *field*.[24] From this perspective, the examination of the material collected by the predecessors and given to the archive could be ignored (Beyer 2011:3).

[21] In the 1930s, the founders of the Ethnological Archives of the Museum of Ethnography in Budapest, for example, followed the Scandinavian model (Forrai 2000:614).The Irish Folklore Commission's archive, which was founded in 1935, explicitly adopted the Uppsala system (Briody 2007:325–331).

[22] Once again these processes have manifested in the folkloristics of different countries with different force and at a different pace, but according to Terry Gunnell, the themes of international congresses and publications clearly show such a trend. Cf. Gunnell 2013:172–173. That is exactly why he felt the need to revisit the anti-archive sentiments in an international forum. At the ISFRN congress in Vilnius, a round-table discussion was organized in 2013, which also appeared in print: Gunnell et al. 2013.

[23] According to Fredrik Skott, the archive structure was not able to follow the changed folklore concepts. See, for example, his contribution to the Vilnius Roundtable on Swedish archives: Gunnell et al. 2013:199–200.

[24] Two other factors contributed to these processes. One is that the subject matter of the scholarship, European folk culture, has completely transformed almost everywhere. The other is that the paradigmatic shifts were supported by the technical innovations of the early 20th century. 20th-century recording techniques (video recording, tape recordings) becoming commonplace in folkloristics has played a major role in the rise to prominence of a narrator-centered approach and the boom of the performance school and contextual trends. Voigt 1997; Kolovos 2004:24.

It was seen that the texts contained in the archive could only be used to research what folkloristics considered folklore in a given period of time.[25]

Although the rejection of the historic-geographic trend and of positivist, context-free data accumulation resulted in a productive rejuvenation of the discipline; the total disregard of its methods and theories was accompanied by the demonizing of the earlier material and collection methodology of predecessors. This, in many ways, brought the archives into a crisis situation. Folklore archivists needed to design new archival systems and structures, as the earlier guises did not meet the new requirements. Due to traditional folk culture transforming at this same time as well, folklore archivists also needed to think about the scope of their collection activites and the social role of their archives. While the latter two tasks were relatively easy to implement in many places, the institutional system could not be transformed as quickly, and the archives could not be easily adapted to the constantly changing questions of researcher(s) and research. What would become of the previous systems also became an issue. Should they be terminated and new ones initiated? Could numerous systems be used simultaneously?[26] The answer to these questions has also been made more difficult by the fact that in many archives, especially in countries where the collections were created in the first thirty years of the 20th century, researchers were faced with the immensity of the systematization and cataloging of collections accumulated in the previous paradigms.[27] Still, it is understandable that for studying folklore and, in particular, orality, interpreted through social context, use, and performance, folklorists required a new documentation process that was more detailed and able to preserve the qualitative features of fieldwork. In Northern European and American folkloristics, researchers who were the most critical of their predecessors were the ones most concerned with these issues.[28] In his writings, Lauri Honko labeled the earlier material of folklore archives as "dead artifacts" that lost their context,[29] while at the same time he and fellow researchers of the University of Turku undertook the

[25] Starting in the 1970s, folkloristics has declared folklore archives and folklore text editions useless because, in the words of Terry Gunnell, "The legends and wonder tales found in the archives and published folk tale collections had all been collected wrongly and for the wrong purpose, as part of an elite-controlled national-romantic agenda, and could only really be looked at from that viewpoint". Gunnell 2013:171.

[26] For the issues that emerged, see the *NIF Newsletter* 1978. 6(1); 1982. 10(4); 1989. 17(4).

[27] In fact, in the Ethnological Archives of the Museum of Ethnography in Budapest, the systematic arrangement of the accumulated material could only commence in the third decade of its operation, in the 1960s. The finalized professional regulation that was necessary for this was only published in 1967 (Forrai 2000:618–619). Sean Ó Súilleabháin, the head of the archives of the Irish Folklore Society, is somewhat skeptical in his 1970 report on the status of cataloging work. In it, he reports that while indices by collector, location of collection and informant are up to date, to prepare the thematic indices of the manuscripts would take roughly six full-time employees doing nothing but cataloging about 20 years of work. In the case of the Irish, cataloging was made more difficult by the fact that only a few people were fluent enough in Irish to be able to process Irish manuscripts. For details on the process and struggles of cataloging, see Briody 2007:325–331.

[28] Cf. the newsletters of NIF (Nordic Institute of Folklore). From 1974 until the 1990s, NIF organized conferences on folklore archiving. Cf. Honko 2001. At the same time, elsewhere the criticisms and disciplines restructured along collection methodology have not led to completely new archiving techniques but rather to a complete distancing from the idea. For the different archival attitudes of anthropologists, folklorists and ethnologists of religion see Mahlamäki 2001:2–3.

[29] Lauri Honko's expression (*dead artifacts*). Cf. Anttonen 2013:159–161; Gunnell et al. 2013:173.

development of a fieldwork methodology in which the data is recorded in accordance with the later archiving system and the recording method contains all the information about the folklore material that is indispensable for its interpretation.[30]

The history of data recording in folkloristics is often embedded in an evolutionary developmental narrative, according to which technology makes collections better, the data more accurate, and consequently, scientific results more credible (Gulyás 2015:24). However, new techniques introduced to capture orality such as audio recordings and video footage, or meticulous, detail-oriented documentation do not actually solve all the issues of the media shift, and the academic study of orality, in turn, continues to require that researchers produce readable texts (Honko 2000a:30). Although there have been many innovative experiments, such as Charles Briggs' book on Mexican verbal arts, where the texts have been recorded in two languages, almost like a musical score, with the various paralinguistic diacritical elements including gestures, intonation, volume, mimics and context (Briggs 1988), they have not been able to substantially transform publication practices. Thus, the performative revolution only confirmed the researchers in two important factors. On the one hand, that a media shift without distortion is impossible, no matter how perfect the technology and how deep the detail in recording the oral presentation; and on the other, that in archival and textological practice, there is no general principle to be laid down and it cannot be fully standardized because it is always the method best suited to the research goal or issue that should be applied (Honko 2000a:29–36; Fine 1984; Finnegan 1992:174–199).[31]

The postmodern critiques of archives and folkloristics, however, brought on a fortunate upswing in the research of the materials of folklore archives. By critically examining the predecessors' work, the rejuvenated studies of the history of the discipline explored the metadiscursive practices of different folkloristic periods.[32] In Hungarian folkloristics, studying the materials of the folklore archive in recent decades – apart from a few comparative studies – focused on understanding the scientific and social-historical contexts of 19th-century folkloristics.[33] Despite the fact that in certain countries folkloristics has productively used a variety of materials from repositories and (non-folklore) archives (witch trials, ecclesiastical visitation protocols, documents

[30] The procedure was named CollCard (collection card) and they started applying it at the University of Turku in the late 1980s. It was already in use when the monumental Siri epic was recorded in India, which later became a three-volume publication. Cf. Honko 1998, 2001; Mahlamaki 2001. CollCard techniques were used to capture, for example, the emic classification of informants during fieldwork and the context of the fieldwork (with/without audience, group, authentic performance, induced context, hidden documentation, active audience, passive audience), etc.: Rajamäki 1989. As a new ideal of documenting, they called for the production of *thick corpora* (Honko 2000b:21–22); elsewhere, the meticulous documentation they used was also called *textual ethnography*.

[31] Therefore, many recommend that a researcher make a simple, legible text that is accompanied by a thick description of the performance/event. E.g., Honko 2000a:36.

[32] E.g., Briggs 1993.

[33] The publication of mainly 19th-century and turn-of-the-century collection manuscripts and related ancillary material (such as correspondence) thanks to the work of Mariann Domokos, Judit Gulyás, Katalin Olosz, Anna Szakál, Imola Küllős and István Rumen Csörsz. For an overview of relevant Hungarian research, see Bárth 2012.

of the Holy See, etc.),[34] folklorists who are receptive to historical topics also rejected folklore archives.[35]

In order for the folklore collections of the late 19th and early 20th centuries to be relevant again, digital technology was required.[36] The database as a special expression of computer culture (Manovich 2009) gave a new impetus to the often neglected folklore archive research and folkloristic textology. In fact, the database once again brought the archives into the focal point of research. This process can be detected in the pursuits of leading folklore societies, where the issue of the legacy of folklore archives and collections is becoming more and more evident.[37] Within both of the most prestigious international ethnographic organizations (ISFNR: International Society for Folk Narrative Research, SIEF: Société Internationale d´Ethnologie et de Folklore), separate working groups and committees address the issue.[38] The theme of the 2009 Summer School organized by the Finnish Folklore Fellows was the relationship between fieldwork and archiving.[39] In June 2015, the Summer School focused on digital folklore, where two separate sections were organized to discuss the issue of digital folklore databases and the editing of digital folklore texts.[40]

In addition to conferences and congresses, a number of periodicals and essay collections have been published on the subject.[41] In the early 2010s, Timothy R. Tangherlini outlined the tasks and challenges of a new research paradigm he named "computational folkloristics" (Abello et al. 2012; Tangherlini 2014). In the last number of years, there are more and more computer analyses of folklore databases available, replacing the earlier studies that were mainly illustrative and merely described the technical details of digitization projects (Tangherlini 2016; Kenna et al. (eds.) 2017).

34 For historical folkloristics in Hungary, see most recently: Bárth 2012.

35 Of course there are always exceptions, such as Lauri Honko's early work, or Anna-Lena Siikala's research, etc. See Harvilahti 2012:405.

36 Although computer methods for the analysis of folklore texts have been used since the 1960s (see Voigt 1981) and the digitization of archive catalogs have been done in many countries since the 1970s (cf. *NIF Newsletter* 1982.10(4)), the paradigm-shifting role and significance of computer technology evolution has only become evident in the past 10 to 15 years with the emergence of the World Wide Web and the widespread adoption of online databases.

37 ISFNR Vilnius 2013, SIEF Zagreb 2015, SIEF Working Group on Archives and the Latvian Folklore Archive, as well as the conference "Towards Digital Folkloristics" in Riga in 2016 organized by the Network of Nordic and Baltic Tradition Archives, and SIEF Göttingen 2017.

38 The idea of setting up a new subcommittee for examining the digital database-based processing of folktales and the spread of tale traditions on the internet first came up at the ISFNR congress held in Tartu in 2005, which was eventually realized in 2009 at the congress in Athens and called Committee for "Folktales and the Internet". See more on this: http://www.isfnr.org/index2.html. At the 2013 SIEF congress also held in Tartu, the SIEF Working Group on Archives was set up, which specializes in the digitization of folklore archives. See http://www.siefhome.org/wg/arch/index.shtml. (accessed June 6, 2017)

39 http://www.folklorefellows.fi/principles-of-fieldwork-and-archiving/. (accessed June 6, 2017)

40 See: http://www.folklorefellows.com/?page_id=2648. (accessed June 6, 2017)

41 For this, cf., for example, the special issue of the journal *Oral Tradition, Archives, Databases and Special Collections* (2013), http://journal.oraltradition.org/. (accessed January 10, 2017); also Holger et al. (eds.) 2014.

DIGITAL FOLKLORE DATABASES – INTERNATIONAL APPROACHES

The purpose of this paper is to provide an overview of the international endeavors of the past 15-20 years in the spirit of the above-mentioned new trend, with a particular focus on how digital folklore databases deal with the folklore archives of the early 20th century and the related theoretical and methodological issues in the new media space; that is, how the material is published in the digital sphere, and along what concepts and methods they intend to give voice to the dead artifacts of earlier collections. It is not an exhaustive overview, but rather an attempt to capture the dominant trends through some of the most important projects, and to highlight the opportunities and challenges inherent in digitization.

Looking at folklore databases, two distinctly different approaches emerge that can be traced back to historical precedents and to the original structures of archives and collections. One is the practice of digital databases based on folklore genres, and the other is the digitization of the entire material of a collector or collection network.

Focus on genre

Genre databases are the continuation of the earliest folkloristic textological practices. The publishing of folktales, ballads, folk songs, and folk legends began with the text editions modeled after the genre hierarchy of 19th-century literature, and peaked in the first half of the 20th century with the catalogs of historic-geographic methods, making genre one of the most important organizing principles of archived folklore materials.[42] Despite the multidirectional criticisms of folklore genres and genre in general, this still remains a characteristic procedure of folklore textology.[43] The popularity of databases focusing on folklore genres is evident in the fact that, while browsing the Internet, one can find, without much effort, for example, databases of Finnish and Estonian runes (Saarinen 2001; Harvilahti 2013)[44], Estonian limericks, folk legends, riddles (Järv 2013:295–296)[45], Icelandic folk legends (Gunnell 2010)[46], Pan-Hispanic ballads,[47] Sephardic Jewish folk poetry (Rosenstock – Bistué 2013)[48], Israeli proverbs (Belinko – Kats 2014)[49],

[42] For the idea of hierarchical classifications following the example of natural sciences, cf. Tangherlini 2013b:39–40.

[43] On the issue of folklore genres up until the 1980s, see Ben-Amos' critical overview: Ben-Amos 1981, later Finnegan 1992:127–147. On. the issue of emic-etic genre categories: Ben-Amos 1969; on the instability of genres: Shuman et al. 2012:61–62. On the recent theories of folklore genres, most recently: Shuman et al. 2012.

[44] http://skvr.fi/ and http://www.folklore.ee/regilaul/andmebaas/. (accessed June 6, 2017).

[45] All Estonian folklore databases can be accessed at the following link: http://en.folklore.ee/dbases/. (accessed January 10, 2017).

[46] http://sagnagrunnur.com/en/. (accessed June 6, 2017).

[47] http://depts.washington.edu/hisprom/, (accessed January 10, 2017).

[48] http://sephardifolklit.illinois.edu/. (accessed April 21, 2017)

[49] Israeli Proverb Index Project (IPIP), currently not available online.

English broadside ballads (Fumerton – Nebeker 2013)[50], or Romanian love charms (Golopenția 1997).[51] In this genre-specific approach, the folktale is clearly in the lead, as there are Dutch, Flemish, Portuguese, Catalan, Armenian, Danish, Icelandic, German and French online folktale databases.[52]

As they are based in folklore genres, these genre-focused databases have also inherited the objectives of historic-geographic method(s) and comparative approaches. One of the primary purposes of their analyses is to re-consider a number of the issues of comparative folkloristics and to answer them in an innovative way using digital technology.[53] What does digital media make possible now that had no solution before? The undisputed advantage of databases is that they are capable of publishing a much larger number of texts than any previous printed collections. While many question the emphasis on quantity (such as what do we gain by analyzing 1 million tales instead of 1,000?), large-scale data analyses or computational literary studies claim that large volumes of digitized content will result in a significant transformation of what consitutes the canon (Moretti 2000; Jockers 2013).[54] In folkloristics, this also means that databases make it possible for less representative, truncated texts, fragments and variants to be included in the research. Thus, one can use not only the pre-selected texts produced by previous generations along textualization processes corresponding to their own era and research objectives, but also a more complete and perhaps less pre-determined corpus (Jockers 2013; Tangherlini – Leonard 2013).

Another great advantage of a digital database is that it allows texts to be ordered in multiple ways, even by combining previous editing practices. One does not have to decide whether to publish the material according to region, settlement, ethnicity, collector or informant. The database allows for switching between the basic data of the folklore material that has been recorded in the database easy and quick. For example, one can search the databases of Icelandic folk legends, Dutch folktales, or Finnish and Estonian runes by collector, informant, or location. For folkloristics, the most liberating novelty of databases is that texts do not need to be assigned to a single category or type. The texts can be assigned to several categories simultaneously, which makes capturing the multidimensional link between them finally possible (Meder 2014a; Abello et al. 2012; Tangherlini 2014; Holger et al. (eds.) 2014). In fact, an appropriate digital textualization not only ensures full-text search capabilities, but also makes it possible through various text mining tools that the multidimensional category system be generated

[50] https://ebba.english.ucsb.edu/. (accessed January 18, 2017). Also the Broadside Ballads Online project at Bodlein Library in Oxford, where English broadside ballads were collected from the early modern era until the 20th century. http://ballads.bodleian.ox.ac.uk/. (accessed January 18, 2017).

[51] http://cds.library.brown.edu/projects/romanianCharms/. (accessed January 18, 2017).

[52] Theo Meder's overview of the online fairy tale databases he knows of. (Meder 2014b:2). http://www.isfnr.org/files/CommitteeInternet.pdf, (accessed June 6, 2017).

[53] For a rethinking of the comparative method and historic-geographic trend(s) independent of digital databases, see Linda Dégh's introduction and other studies in the special edition of the *Journal of Folklore Research,* "The Comparative Method in Folklore." Dégh 1986, and Virtanen 1993; Wolf-Knuts 2000. Most recently it was Frog who presented in his paper how historic-geographic methods can be applied productively in contemporary folkloristics. Frog 2013:23–30.

[54] For more on the theoretical assumptions underlying the problem, see "Theoretical frameworks" of the current study.

and determined not only by the researcher(s) but also by the algorithms that can analyze the texts to find characteristics that show similarities among certain texts.[55]

Digital technology also solves one of the highly criticized points of mapping. In the past, an insurmountable difficulty of analog ethnographic atlases was the rigid representation of diachronous data on a synchronous map, the very reason for most of the criticisms of the cartographic method (Munk – Jensen 2014:40–41).[56] But digital maps can represent the dimension of time and space simultaneously. In the databases of Dutch folktales and Icelandic folk legends, one can choose which period's texts one wants to see on the map.

The new medium thus aids and reforms folklore studies in many ways, and we can expect many results from it. At the same time, the concept of the genre-based database raises several theoretical and methodological issues. It seems that the corpora are practically given for genre databases, or at least that is what the archives' genre-based card indexes and genre-based folklore text editions seem to suggest. In fact, some databases are merely a digitized version of a previously closed corpus. Such is, for example, the Finnish rune database that made the 34-volume *Suomen Kansan Vanhat Runot* (The Ancient Poems of the Finnish People, 1908–1948, 1998) available, thereby making more than 89,000 texts digitally searchable.[57]

A more complicated case is when the authors create a database by selecting the most important volumes representing the genre. An example of this is the Icelandic folk legend database, the *Sagragrunnur*. In the past, Icelandic folk legends did not have an index, collection or catalog of any sort, and publications did not reference international parallels or types. Terry Gunnell and more than 25 of his students published a digital database that referenced the text of about 10,000 Icelandic folk legends. It is important, however, that, contrary to the databases mentioned earlier, texts are not published here; in fact, the only one available is the Icelandic, and in some cases English-language, abstract of the texts, with an English-language search interface and keyword search option, which accomplished the innovative replacement of a never-existing catalog of Icelandic folk legends, thus completely bypassing the textualization problems arising from various publications (Gunnell 2010).

Although previously published books and catalogs can be relatively helpful in selecting texts that should be part of a collection in a given genre database, a variety of dilemmas may arise around the building of a corpus, as different cultures and eras, as well as folkloristic practices, feature different genre repertoires, and databases usually select texts based on a genre concept that was constructed and solidified in the late 19th, early 20th century, despite the fact that they cover an extended period of time. Who decides whether a 17th-century religious-magical text is a prayer or an incantation? Another complicating factor is that in order to include as many texts of the given genre as possible, genre databases usually go beyond the limits of physically existing

[55] For more on text-mining methods, see "Text mining and network theory methods" of the current study.

[56] For more on digital maps, see "Visualization" of the current study.

[57] However, the Estonian counterpart of the database called *Eesti regilaulude andmebaas* (The Database of Estonian Oral Poetry), which basically follows the Finnish model and contains 80,000 texts, also includes manuscripts and published texts; moreover, if a text exists in two versions (manuscript/publication), one can switch between the two transcripts.

text collection(s) or archives.[58] Thus, databases need to display and manage a fairly heterogeneous source type. The Dutch folktale database, for example, includes texts from 16th-century manuscripts, 20th-century folklore collections and newspapers, as well as the Internet (Meder 2014a).

Another criticism could be that databases based on artificial genres divide oral culture and are therefore not suitable for comprehending the complexity of verbal arts or for examining the texts of a given locality, and are, in fact, and impediment to re-imagined comparative folklore studies, as independent databases do not show the multifaceted relationship of motifs and themes across genres. A good example of this is the case of Estonian folklorists who have tapped the potential of the digital technology from the very beginning and now have more than twenty types of genre databases. At the moment, however, one of their biggest challenges is how to integrate these genre databases into a joint digital repository, thus ensuring joint search and research capabilities across a variety of virtual archives (Kõiva 2003; Järv – Sarv 2014:55–56; Järv 2013:295–296). Recognizing the issue, there is currently a growing number of thematic folklore databases that are operating in comprehensive genre categories. The Dutch folktale database, for example, retains its folktale nature only in its name. The extended purpose of Theo Meder and his colleagues was to document the entire stock of Dutch folk prose, which is why the *Nederlandse Volksverhalenbank* stores many kinds of material, from anecdotes and myths through folk legends to personal stories (Meder 2014a). The Norwegian magic database of the University of Oslo has been compiled by processing three types of sources – early modern grimoires and witch trials, as well as folk legends collected in the 19th and early 20th centuries – which are published on a common interface.[59] The Portuguese legend database includes sacred, historical and urban legends, origin myths and supernatural stories.[60]

With such degree of heterogeneity, it is questionable how databases can capture and display the context of these diverse sources, and how they can solve the textualization issue raised by the different types of texts.[61]

58 The dilemma of corpus building is characteristic of the digital humanities in general; a Hungarian literary example: Gergely Labádi's study of the compilation of the 19th-century corpus of the "Hungarian novel". Labádi 2014.

59 Trolldomsarkivet, Norsk Folkeminnesamling. http://www.hf.uio.no/ikos/english/services/norwegian-folklore/magic-in-norway/ The database can only be searched and used in Norwegian: http://www.edd.uio.no/ikos/trolldom.html (accessed June 20, 2017)

60 http://www.lendarium.org/ (accessed June 20, 2017).

61 For the issue, cf. "Technique. Platform independence, interoperability, sustainability" of the current study.

Focus on the collection

The other main approach of folklore databases is to digitize the entire collection of a prominent collector, which in the case of European archives often coincides with the important foundation or key collection of the institution[62] and therefore enjoys priority due to its antiquity.[63] This approach can be justified not only by the digital preservation of physical objects, but also, as indicated at the beginning of the study, by the fact that ancillary materials linked to research questions make up a major part of the folklore archives, and in such cases, the objective must be the digitization of this complex material. I would like to compare two important examples. The digitized collection of Danish folklorist Evald Tang Kristensen (1843–1929), *Danish Folklore Nexus* and *ETKSpace*, by Timothy R. Tangherlini, and the *WossidDia*, i.e., the *Digital Wossidlo Archive*, which, made the collection of Mecklenburg folklorist Richard Wossidlo (1859–1939) available digitally under the leadership of Christoph Schmitt.[64] The most striking similarity between the two collections that were created around the same time is their almost inconceivable volume. During almost fifty years, Kristensen recorded more than a quarter of a million texts (including ballads, folk songs, folktales, folk legends, proverbs, incantations, limericks/rhymes, folk games, jokes, folk medicine, and other descriptions of everyday life) that were collected from about 4,000 informants. Altogether, the collection contains 12,000 site names, 60,000 collection data, and more than 24,000 pages of manuscripts (Abello et al. 2012:63–65; Tangherlini 2013b). During his ethnographic and dialectological collections, Wossidlo also focused on a great deal of subjects such as folk customs and beliefs, coastal rural labor and everyday life, ethnobotanics and sexuality, and genres such as the folktale, joke, anecdote, and proverb. The Wossidlo Archive of the Institute for Volkskunde/European Ethnology at University of Rostock consists of approximately 2 million data and thousands of manuscripts (Holger et al. 2014: 64–65).

Despite the similarities, the nature of the two collections and the purposes of digitization have resulted in different implementations. Beyond the incredible volume of his collection, Kristensen also stands out from well-known 19th-century folklorists thanks to his high degree of precision including documenting variants, informants' living conditions, life histories and everyday life. His collection is unmatched in its many source types, as it preserved his own manuscripts, drafts, thousands of pages of field notes, memoirs, correspondence with attachments, that is, a multitude of manuscripts of local folklore collections. If we add to this the works he edited and published and the thematic indexes, a picture of late 19th-, early 20th-century collection and editing methodology and

[62] There are also instances of scattered collections being virtually consolidated in an online database (cf. the projects to publish digitally the collections of Adolf Spamer or the Opies, see Seifert – Keller 2014; Bishop 2013). When digitizing the collection of Iona and Peter Opie (project name: Childhoods and Play), multiple types of documents (handwritten and typed notes, questionnaires, correspondence, sound recordings, photographs, newspaper clippings) created between 1950 and 2000 were processed; most of the material is in the Bodleian Library in Oxford, a smaller portion in the Folklore Society Archive in London, and sound recordings are physically located in the British Library Sound Archive. http://www.opieproject.group.shef.ac.uk/. (accessed January 18, 2017). http://www.opieproject.group.shef.ac.uk/about-collection.html. (accessed January 18, 2017) (Bishop 2013:205–206).

[63] The collection of Hungarian folklorist, Lajos Kálmány was advanced for this reason, Granasztói 2008.

[64] http://etkspace.scandinavian.ucla.edu/macroscope.html, https://apps.wossidia.de/webapp/run (accessed June 27, 2017).

textualization practices emerges. The Danish Folklore Nexus offers a glimpse into only one, somewhat small, corpus, compiling about 500 texts from five informants, but it does so in their full complexity (Tangherlini – Broadwell 2014). Its greatest merit is that we can view simultaneously the field-collection trips, and the different texts they produced. On a computer screen, the scanned image of the manuscript can be placed side by side with the translation of the manuscript, and we can even display the version edited and published by Kristensen. Each field-collection trip can be traced on a digital map (both 19th-century and modern), making it possible to connect the research trips and the collected materials and their social contexts. One can quickly switch perspectives at any point between the field sites, informants and their repertoires, even the collector's manuscript.

Because Kristensen's collection is important not only for its complexity but also for the enormous amount of material, another interface, ETKSpace, was created to handle it.

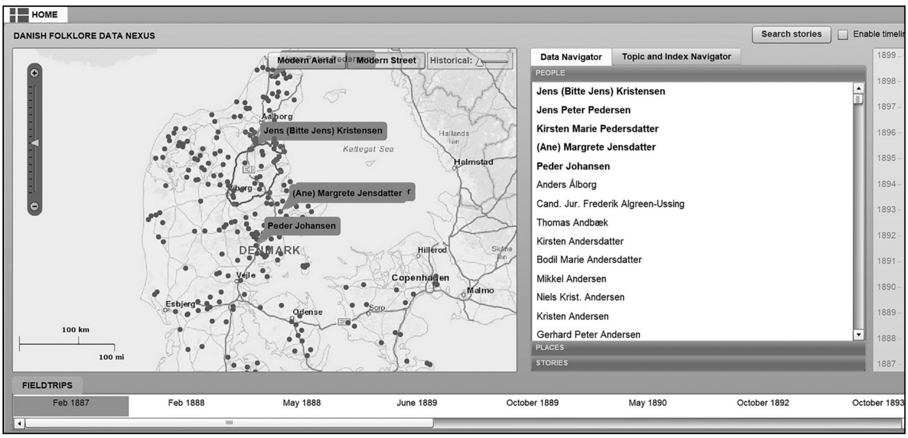

Figure 1. Danish Folklore Nexus, the 1887 collection trip of Evald Tang Kristensen. http://etkspace.scandinavian.ucla.edu/danishfolklore/# (accessed September 28, 2017)

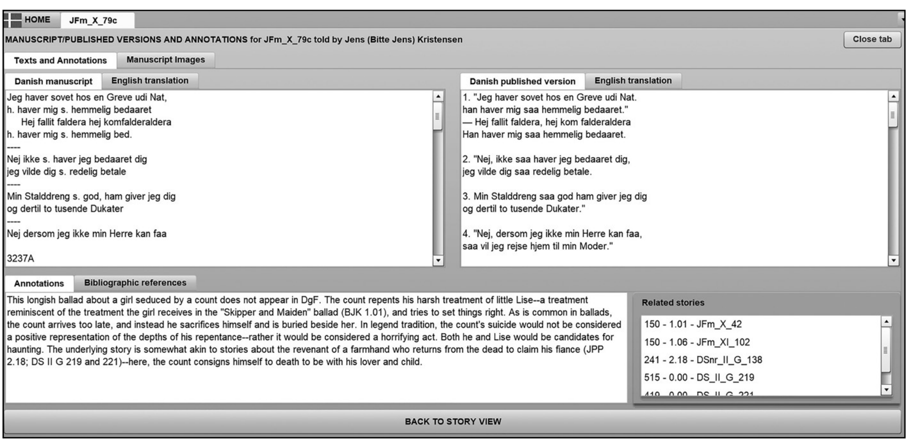

Figure 2. Danish Folklore Nexus, window system for parallel reading of different text versions. http://etkspace.scandinavian.ucla.edu/danishfolklore/# (accessed September 28, 2017)

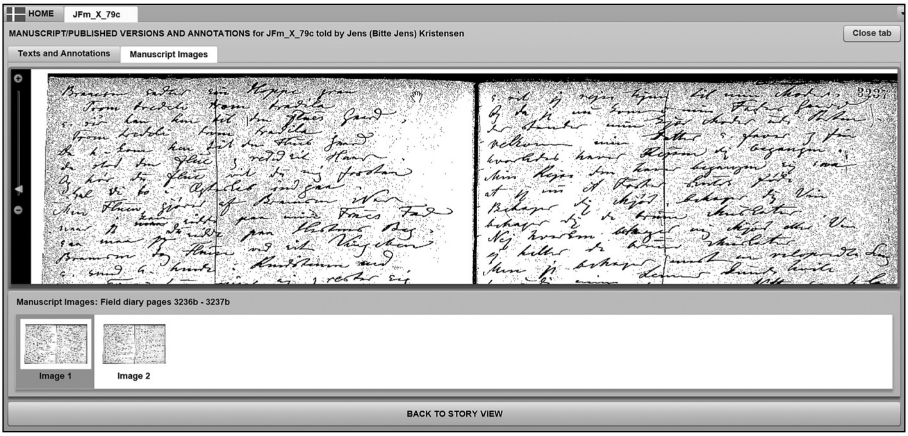

Figure 3. Danish Folklore Nexus, scan of the original document. http://etkspace.scandinavian.ucla.edu/danishfolklore/# (accessed September 28, 2017)

Here, in the digitized version of the full, published material, search capabilities and access to texts is ensured, but we have fewer metadata to choose from. In a corpus containing more than 30,000 texts, one can search by Kristensen's own indexes, locations as well as a Tangherlini's category system, keywords that encapsulate his earlier research results. On the one hand, the Tangherlini concept created an interface suitable for exploring the collection's context, the different relationships between collection participants and texts, and the textological practices of the folkloristics of the era; on the other hand, the automated digitization (OCR) of the immense volume of Kristensen's printed works made the entire corpus accessible. So the interfaces they created fit well with the new direction of critical archival folkloristics mentioned at the beginning of the paper, while the coherence of the vast amount of material, in contrast to many genre databases, makes it suitable for both comparative research and the contextual understanding of 19th-century narratives of rural Denmark. Apart from the aforementioned positive aspects, one criticism of the Danish Nexus and ETKSpace can be the fact that the sources of the collections are not sufficiently displayed on the digital interface. The website does not provide an introduction or information about what resources were consulted, or what percentage of the corpus has been made available and how. Users of the website may figure out what the the abbreviations used in the database stand for with the help of Tangherlini's printed book (Tangherlini 2013b).

In contrast to the above, WossiDia attempts to display Richard Wossidlo's material in the context of the archive, in the complexity of the archival system. The largest part of WossiDia consists of the handwritten notes of Richard Wossidlo, which mainly document the Mecklenburg collections he or his collectors acquired. Wossidlo made notes not only on fieldwork data but also on Mecklenburg data gained from various published sources such as journals, books and newspapers, and their parallels in German and, to a lesser extent, Slavic and Scandinavian ethnography. The second most important unit of the archive is the collector's corpus of correspondence. However, the true significance of the collection lies in the archival system itself, which, with its various cross-references, attempted to document the semiotics of a rich and diverse cultural system. In the first phase of digitization, the entire collection was scanned including all notes and letters, and then a complex directed

Figure 4. WossiDiA, the original index card catalog. https://apps.wossidia.de/webapp/run (accessed September 28, 2017)

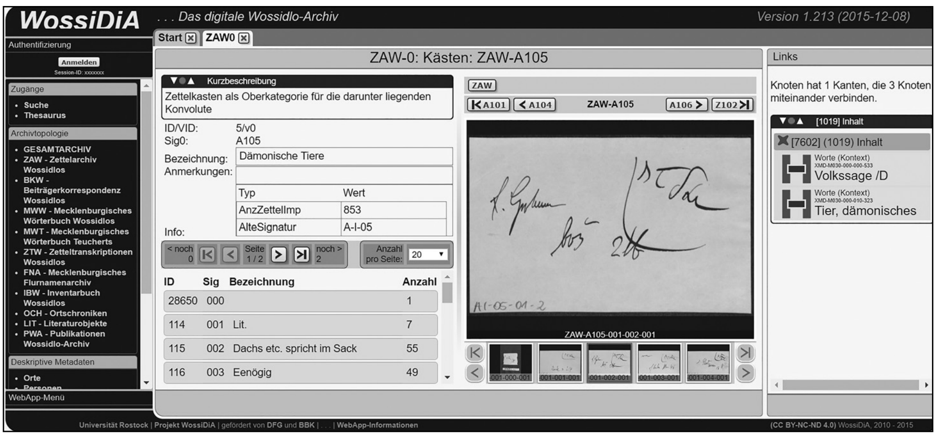

Figure 5. WossiDiA, visualization of a digitized index card in the database. https://apps.wossidia.de/webapp/run (accessed September 28, 2017)

hypergraph model and search interface was created which was capable of simultaneously storing and displaying the possible connection points between different documents. The goal of the creators was to "re-contextualize the lost link between researchers and collectors and their networks within an intelligent digital archive" (Holger et al. 2014:84).

According to the creators of the database, even though the WossiDia was designed specifically for the Wossidlo archive, the model they developed can generally be used to digitally process and display complex ethnographic collections. It should be added, however, that WossiDia was able to do this because they were working with Wossidlo's existing rich category system and cross-references. A theoretically and practically closed and completed archival system was transposed into the digital sphere, thus discovering links between elements that would not have been possible during a manual search. At the same time, in this approach, digital edition does not override the existing category system, the indexes and descriptors, so the foundation set by Wossidlo ultimately determines the database, too, along with its possible errors. Most folklore archives are not so lucky. As mentioned previously, due to their analog nature, archives could never complete the organization of their material, but a prerequisite of a solution like WossiDia is a well-structured, well-organized folklore material.

One of the main objectives of digital archives and electronic text editions is to publish data and texts in their original context (McGann 2010). What does this mean, however, from the perspective of folkloristics? Obviously, folklore texts can never be restored to their original, primary context of utterance, but these texts have many other contexts (Anttonen 2013), which is exactly what is exciting and complex about database creators trying to determine what contexts they should restore the data to reflect. Genre databases placed texts in the context of space and time and many millions of similar texts; the Danish Nexus placed them in the context of the historical and geographical relationship between collection, collector, informant and 19th-century Denmark; while WossiDia contextualized its material within the archival structure and the collection. However, their digital mode also places them in a new context, in the world of digital databases, where the focus is on interoperability and long-term sustainability. In view of this, it is also necessary to talk about a third relevant issue, namely, how genre- and collection-centered databases can be integrated into the digitization concepts of complete folklore archives.

Focus on the archive

A common issue of genre- and collection-centered folklore databases is the creation of categories necessary for data retrieval, an issue that is no longer independent of how folklore databases operating on different concepts can be integrated into the complete digitization project of certain folklore archives.[65] How and where do they connect, how can they be linked together? My study cannot go into detail about the general question of the digitization of folklore archives, as it focuses on folklore databases, but some things need to be addressed, mainly because folklore databases are typically part of a folklore archive and their sustainability can only be ensured by these institutions.

According to Jerome McGann, who mainly studies 19th-century English literature, one of the great lessons of archival history is that mankind always fails in trying to decide what is worth preserving. Many believe the illusion that this almost entirely digitized culture fully maps the physical archive, which is why it is unnecessary to visit it anymore (McGann 2014:42–45). Although digitizing and publishing material in its entirety is in many respects not the objective of folklore archives, due to ethical and legal issues and lack of resources, McGann's ideas are worth considering. What needs to be digitized? The earliest? The most complete? The typical, the individual, or the disintegrating? In many cases, the physical maintenance of the objects may be

[65] On the digitization of Estonian and Irish folklore archives, see Järv – Sarv 2014; Ógáin 2013. In the Hungarian context, on the development of the Ethnological Archives of the Museum of Ethnography in Budapest see Granasztói 2013.

neglected on account of digital preservation that has an uncertain lifespan, even though the digital image is merely a numerical code that is a thousand times more vulnerable than 19th-century paper or ancient parchment. Seeing the two million scanned notes in WossiDia, one might rightfully ask: what is the principal task of the digitization of an archive or collection? Is it worth digitizing the often erroneous notes with a few lines of references just so someone could have access to them from home? Or should a folklore archive focus on revealing the content of the material as accurately as possible, providing digital aids and indexes to ensure the most detailed data retrieval (Granasztói 2008:126–127)? At the same time, because of the aforementioned disorganization of folklore archives, in some cases this can only be achieved with the full digitization of the content, as we will see below.

DIGITAL CONTENT CREATION

After a brief history of folklore archiving and a concise summary of digital folklore databases, let us now see how these concepts can be implemented in practice. What can digital humanities offer, what new issues are emerging, and how have folklorists implemented these ideas so far?

Practical issues and resources

One of the most important steps in digitization efforts is the creation of machine-readable textual data. How do we create this resource? How can we give voice to the millions of texts in the archive that have not yet been cataloged? The digital transcription of manuscripts is expensive and time-consuming, and folklore archives that typically work with but a few specialists are unlikely to be able to carry it out on their own. As a solution, methods like *citizen science* or *crowd science, crowdsourcing, civic science* (volunteer-based science, community collaboration)[66] have become popular. *Crowdsourcing* involves volunteer civilians in data production or even in certain phases of the data collection process. Volunteering is already a familiar notion in folkloristics thanks to the networks of collectors, and it seems that this new version can also be operated effectively. When recording folklore texts, the greatest need for civil volunteers is where OCRs do not work, for example with handwritten documents, or if they do work, they need a continuous manual review such as when typography or language features cause the program to make mistakes.

The National Folklore Collection's *The Schools' Collection* project in Ireland is a good example of the productive application of crowdsourcing. The Schools' Collection was collected by about 50,000 students from more than 5,000 elementary schools between 1937 and 1939, containing approximately 740,000 pages of manuscripts

[66] http://www.oszk.hu/civic. (accessed January 9, 2017). For a critique on his role in digital text editions, see: Shillingsburg 2016; for an overview of opportunities for civic science and digital humanities, see Terras 2016, for a general summary and historical presentation: Riesch 2015.

(Briody 2007:260–270).[67] With the *Meitheal Dúchas.ie: Community Transcription* project, launched in 2013 and operated by the folklore archive at the University College Dublin, in a narrow three-year period, volunteers transcribed over 12,000 pages of Irish and 27,000 pages of English-language material which constitutes 15% of the Irish and 9% of the English-language segment of the entire collection.[68] A similar campaign was also used by the Latvian folklore archives (*Latviešu folkloras krātuve*) for their *Simtgades burtnieki* (Wizards of Centenary) program,[69] where they joined the volunteer civilian preparations for the country's centenary celebrations.[70] In June 2016, the archives launched the opportunity of transcribing their manuscripts to the public, and in almost six months, nearly 30,000 pages of text were transcribed by civilians. In the multi-lingual collection registered users could choose from eleven languages; Latvian, Latgalian, Livonian, Lithuanian, Estonian, Russian, Belarusian, Yiddish, Romany, Polish and German. On the digital interface of the Latvian folklore archive, volunteering works regardless: committed enthusiasts of community science can even transcribe audio materials here.[71]

Obviously, the above cannot be applied in all cases, as some philological pre-qualification is needed to transcribe the handwriting of the 19th or earlier centuries. However, a large part of the immense corpora of European folklore archives created in the 20th century consist of mostly legible handwritten or typed texts of school children, pupils, teachers or rural intellectuals.[72]

The transcriptions of texts must be reviewed by folklorists or philologists proficient in folkloristic textology. This is the case with the Irish and Latvian material. There is still an enormous potential in crowdsourcing, which of course can be used for more than just transcribing texts, for example for data collection, data review, photo recognition, localization of data. In the Latvian campaign, for example, beyond transcription, volunteers are also allowed to add any keywords they create to the transcribed text. Thus, crowdsourcing is not merely a source of free workforce for public institutions struggling to process data. With the involvement of civilians, museums and manuscript libraries can increase their social function and usefulness, leaving the ivory tower of science and interactively making the public become participants in knowledge production. Increased civilian interest in national, folk, local and various cultural traditions, and the former collection networks' capital can, in the case of folklore archives, make participatory

[67] http://www.duchas.ie/en/info/cbe. (accessed January 18, 2017).

[68] A scientific overview of the motivation, speed, and effectiveness of lay participants has also been produced in *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/. (accessed January 9, 2017). The digital dissertation provides a generous bibliography and references of portals that use *crowdsourcing* and scientific writings on the phenomenon.

[69] http://lv100.garamantas.lv/en. (accessed January 9, 2017).

[70] For Latvia's 2018 centenary programs, see: http://www.latvia.eu/latvias-centenary. (accessed January 9, 2017).

[71] http://garamantas.lv/en. (accessed January 9, 2017).

[72] Encouraged by the above example, the same method could be productively applied in the processing of the Hungarian Folklore Fellows manuscript material of more than 10,000 pages found in the Ethnological Archives of the Museum of Ethnography in Budapest, since the texts were noted down between 1912 and 1921 by schoolchildren in an often legible, cursive style.

civic science initiatives particularly viable, which may lead to the revival of the, often never-ending, traditional, voluntary ethnographic collecting.[73]

Technique. Platform independence, interoperability, sustainability

Data production is a time-consuming and tedious process, and still one of the most expensive operations of digitization even if in some cases crowdsourcing initiatives can reduce the costs. For this reason, long-term preservation and widespread utilization are primary requirements, but this is where the most uncertainty can be detected in the paradigm of digital humanities. Which format will hold up in ten years? What should the data be prepared for at all? How and where will the various humanities disciplines find the common denominator in digital textology and digital data preparation to ensure interoperability?

The purpose of digital textology is to mark up all internal and external information associated with the text that could be relevant to its analysis in the future in a machine-readable way. The greatest theoretical and methodological paradox of digital text or data preparation lies precisely in this. Data must be prepared in a versatile way so that it is suitable for as many different types of analyses as possible and reinstatable into as many contexts as possible, and in the virtual world of databases, they should be able to connect to and network with as much data as possible. At the same time, it is not possible to standardize data preparation because it always depends on the purpose of the research, which in digital humanities is often difficult to foresee (Jockers – Underwood 2016:299–300; McGann 2016).

The selected markup language must be simultaneously uniform yet flexible and expandable. To do so, one of the best choices currently appears to be encoding texts in Extensible Markup Language (XML), which is platform-independent and relatively resistant to technical changes. The Text Encoding Initiative (TEI) itself, which provides text encoding suggestions for digital scientific text editions, also recommends the XML markup language because it does not define a finite set of elements but merely sets out the rules for creating a well-formatted markup language.[74] However, the appropriately selected markup language is only one side of the coin; the next problematic question is what material should be marked up and how (McGann 2016).

In folkloristics, the previously mentioned Finnish and Estonian rune databases and the Romanian love charm database were the first to use the XML markup language (Saarinen 2001); ETKSpace and the recently launched Irish The Schools' Collection project followed in their footsteps, but due to their different objectives, the coding procedures of all of these databases are different from each other. Even if TEI is used by folklore databases,

[73] For example, Carsten Bregenhøy reported in his brief presentation on the history of the Danish folklore archive that Danes already lost interest in traditional folklore materials in the 1970s, and raised the question of how to make people interested in maintaining these institutions and where to find a new relationship with audiences. *NIF Newsletter* 1978. 6. In Hungary, the activity of the volunteer collecting network reached its peak in the 1960s-1970s (with around 1,200 applicants), but at the same time professionals found it inexhaustible, as they could not be used for the current scientific purposes (Forrai 2000:632).

[74] For more details on TEI, see most recently: Pierazzo 2016.

```
This XML file does not appear to have any style information associated with it. The document tree is shown below.

▼<xml>
   <metadata type="Keywords">vætte</metadata>
   <metadata type="Keywords">rumsteren</metadata>
   <metadata type="Places Mentioned">Odderup</metadata>
   <metadata type="Story">DS_01_0_00011</metadata>
   <metadata type="ETK Index">Mound dwellers (Hidden folk)</metadata>
 ▼<metadata type="Story Text">
     11. På en gård i Odder up er der sådan en rumsteren på oftet med alle ting på et bestemt sted og en bestemt tid ; men kun hvem der er fodt på höjtidsaftener kan se, hvad
     det er, og de siger, at det er nogle ganske små vætter.  F. L. Grundtvig.
   </metadata>
 </xml>
```


Figure 6. Transcript of a belief legend on ETKSpace encoded with XML markup language. http://etkspace.scandinavian.ucla.edu/etkSpace/export/export.php?col=Story&ID=1&stylesheet=XML (accessed September 28, 2017)


```
<?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>tüübinimed kontrollimata</TYP_KONT>
<TYP_YHT>Sirbiviskamise laul</TYP_YHT>
<LLIIK_YHT>Töölaulud </LLIIK_YHT>

</META>
<TEXT>
<V>Sirise, sirise, sirpi,</V>
<V>Kőrise, kőrise rauda,</V>
<V>Kelle sirpi ette jőuab,</V>
<V>Selle juurde peigu jőuab,</V>
<V>Sellel peigu pőlle ostab,</V>
<V>Ja tasuks tanugi ostab,</V>
<V>Passib pähä naise mütsi,</V>
<V>Naise mütsi, pruudi paelad.</V>
</TEXT></ITEM>

</KOGUM>
```


Figure 7. Transcript of a rune text in Eesti regilaulude andmebaas (The Database of Estonian Oral Poetry) encoded with XML markup language. http://www.folklore.ee/regilaul/andmebaas/?op=1&oid=4 (accessed September 28, 2017)

they do not necessarily find the same elements worthy of tagging. The XML encoding of one of ETKSpace's belief legends reveals what the creators actually coded. We can see the keywords that have been assigned to the text as metadata ("Keywords"), the place names occurring in the text ("Places Mentioned"), Kristensen's typology ("ETK Index"), the text identifier ("Story"), and finally the text itself ("Story Text"). In comparison, in the case of a text of the Estonian rune database, they tagged the location of the collection, information about the collector and informant such as age, and even the line breaks in the part where it appeared in the actual text.

For each transcribed text on the interface of The Schools' Collection, one can download up to three XML files. One applies to the school, one to the relevant manuscript page, and one to the text itself. These three examples are a good illustration of the fact that even with the same markup language applied across textualization processes, text editions with different depths and detail may be produced. The difference may be even larger if a folkloristically important genre is published by representatives of

```
</teiHeader>
<text>
 <body>
  <div type="ballad">
   <div type="part" n="1">
    <head>
     <title>
      <seg n="1" rend="left">
        <hi rend="italic">Witchcraft discovered and punished.</hi>
      </seg>
      <seg n="2" rend="left">
       <hi rend="italic">
         Or, the Tryals and Condemnation of three Notorious Witches, who were Tryed
       </hi>
      </seg>
      <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 Sentance
       </hi>
      </seg>
      <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/>
      <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>
```

```
          <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>
    <closer>
     <seg n="1" rend="left">
       <hi rend="italic">FINIS.</hi>
     </seg>
    </closer>
   </div>
  </body>
 </text>
```


Figure 8a-8b. Transcript of a broadside ballad in EBBA (English Broadside Ballad Archive) encoded with XML markup language. https://ebba.english.ucsb.edu/ballad/31034/ebba-xml-31034 (accessed September 28, 2017)

other disciplines. *The English Broadside Ballad Archive* (EBBA), which has more of a literary and book-historical approach, also uses TEI tags. The encoding of these broadside ballads, however, includes the precise description of the early modern printed matter; author, publication date and publisher, for example, the poetic characteristics of the individual ballads such as stanzas, rows, chorus, and even the typographical features of the edition, whether it was in cursive or in capitals, which could, in theory, accurately reproduce the printed matter in case the object was destroyed, as all of its existing physical qualities have been encoded.[75]

The standardization of digital textology across disciplines is obviously a utopian daydream or nightmare, and uniformization would practically destroy the texts and their pluralistic readings created through various textualizations. The editions should be primarily based on the nature of the material and provide answers to certain questions, even if these questions are multiplied in the digital world. At the same time, another opinion holds that if everyone annotates something different, the corpora may become fragmented. According to Martin Wynne, a leading expert on digital humanities at Oxford University, through a lengthy and meticulous data preparation, we are effectively focusing on future interpretative possibilities rather than interpreting the actual data, and this can only be eliminated by developing better and faster tools for automatic annotation. In this case, the programs would index the corpora through various digital procedures (Wynne 2012).[76]

Despite the above uncertainties, it is important that folkloristics also develop a suitable annotation strategy for its discipline, because if it does not, it may lose properties that are crucial. The TEI, for example, has primarily literary-linguistic motivations and is thus using a philological approach. However, in the case of a folkloristic edition, the tagging of paragraphs is not that relevant, unless we want to produce a historical or textological study, but indicating whether the title or the genre definition of the folklore

[75] http://ebba.english.ucsb.edu/page/tei-xml (accessed January 9, 2017)

[76] For the various resolutions on the issue, see also: Pierazzo 2016:316–318.

text is etic or emic, whether the given notation was made from memory, shorthand or a sound recording, and whether the collection was the result of an informal interview, questionnaire or participant observation would be justified, even if in many cases these are not known. In the world of digital text editions, the textualization of the source text has replaced the production of the main text.[77] The catalog systems of historic-geographic method(s) and the notes of folklore archives still fulfill this role, but the later, oral history-type folklore materials, field notes, and other ancillary notes can no longer be objectified and divided so easily, and the descriptive metadata templates describing physical objects do not favor the digital archiving of intangible heritage and orality.

There is much debate about annotation, but it is clear that true interoperability between different databases and corpora can only be implemented if database creators and designers annotate the same properties of the texts and, preferably, in the same system, or at least in compatible systems.[78] Consistency and interoperability between databases are key, for which there is no strong example in folkloristics to date.[79] The development of international standards in folkloristics will take some time, although several people have underlined the need for it recently.[80]

77 According to Attila Debreczeni, a source text is a manuscript or printed document of physical existence. For more on switching between main text and source text, See: Debreczeni 2014:32.

78 There are countless initiatives within digital humanities (e.g., Dublin Core Metadata Harvesting: DCMI, Open Archives Initiative Protocol for Metadata Harvesting: OAI-PMH, CLARIN) to standardize these metadata in hopes of an exchange. Meder 2014a:125. On a proposal for the standardization of digital textualization procedures of different national charm and incantation databases, cf. Ilyefalvi 2017.

79 The first such project might be *ISEBEL: Intelligent Search Engine for Belief Legends*, whose search engine will be able to simultaneously analyze texts found in the Dutch, Danish and the Mecklenburg folklore databases. At a panel on archives at the SIEF congress in Göttingen (2017), Theo Meder announced that a grant made it possible for them to co-operate with Timothy R. Tangherlini and Christoph Schmitt to coordinate the three databases and develop a search engine. See also https://www.researchgate.net/project/Trans-Atlantic-Digging-into-Data (accessed June 20, 2017). From the perspective of integration, it is also relevant for folklore archives how they can connect to the digital archiving standards of other national/international institutions and what the consequences, advantages and disadvantages of that might be. Well-thought-out joint projects can greatly help and simplify digitization. The scanning of the important manuscripts of the Estonian folklore archive (for example, of founder Jakob Hurt), is happening through the uniform digitization of the Estonian Literary Museum that maintains it. It is therefore not necessary to include manuscript versions in image format in the aforementioned Estonian rune database, as they can be displayed very quickly in the other database, and it is obviously only a matter of time when this can be accomplished with direct built-in links without a separate manual search. Another good example of cooperation between institutions and projects is the Icelandic folk legend database. Although one cannot read full texts in the Sagragrunnur, in cases where a volume has already been digitized by the National and University Library of Iceland, a built-in link can immediately help one navigate to the cited page of the volume so that the text can be read in its entirety. www.Bækur.is. (accessed June 21, 2017).

80 Furthermore, at the 2015 workshop of the SIEF Working Group on Archives in Zagreb, the plan for a guide to standardizing the recording of the metadata of folklore databases was raised. E.g.: Meder 2014a:124; Katajamäki – Lukin 2013:13. For the objectives of the SIEF Working Group on Archives, see http://www.siefhome.org/wg/arch/ (accessed June 21, 2017).

Methodological and theoretical issues. Reliability, credibility.

Choosing the markup language and tags is a critical and controversial point of the digitization process because it is not just a technical matter. The important issues that are crucial for the edition are practically determined here, as this determines how the digital data will be used later, and it is difficult to correct it later (McGann 2016; Pierazzo 2016; Debreczeni 2014:2–8). Already at the developmental stage, the creators of the database must be aware of what they want to use the database for, but digital databases are also designed to be scalable and complementary. In comparison with previous scientific practices, constant renewal and instability creates uncertainty. Book-based critical editions of texts have never been perfect either, but at the time of their creation, from theoretical, methodological and practical aspects, they were closed creations and that is how they came into the world of scientific research (McGann 2014:26). Digital databases will, by their very nature, never achieve this. However, in terms of reliability and credibility, it is even more confusing that most digital text editions do not explain what textualization procedures have been performed on their texts, i.e., they do not disclose the "history" of the texts and do not make them available in their raw form. They do not display the entire corpus, the criteria for text selection are not clear, and neither

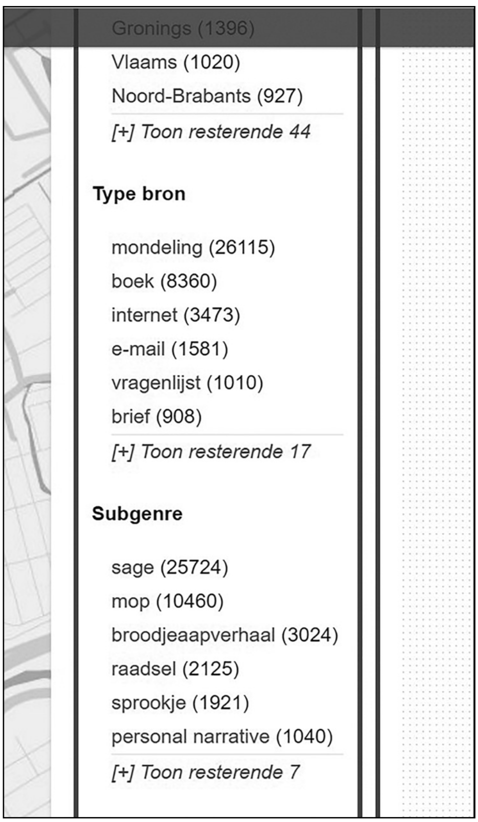

Figure 9a.

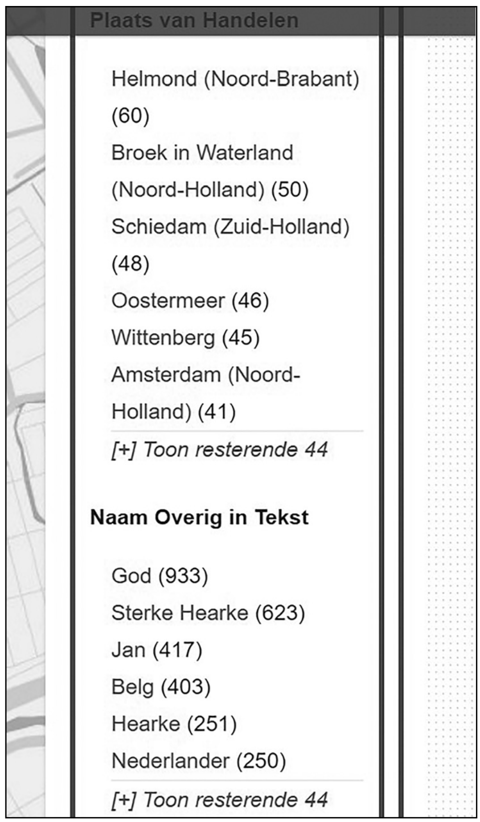

Figure 9b.

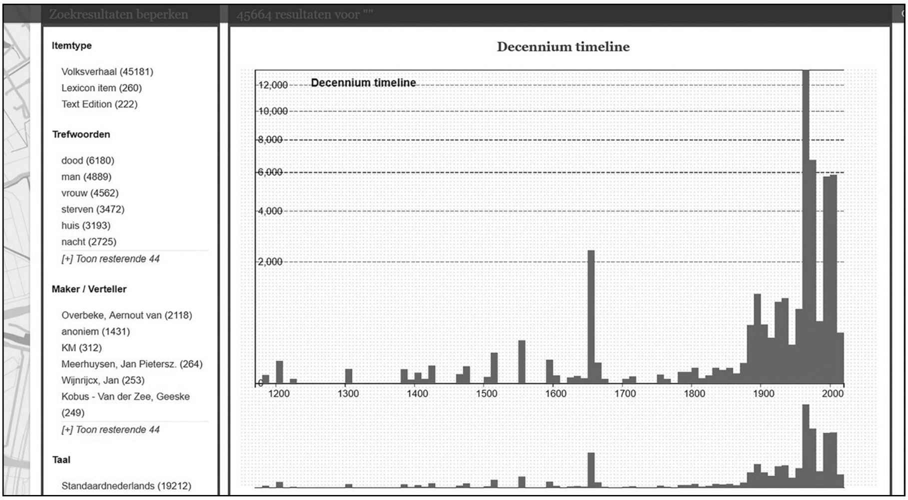

Figure 9c.

Figure 9a-9b-9c. Details from the left sidebar of the Dutch folktale database. Figure 9a shows the quantification of source types, Figure 9b the quantification of word occurrence, and Figure 9c is the dynamic timeline of the folktale database. http://www.verhalenbank.nl/visuals/timeline?q=&facet=&free= (accessed September 28, 2017)

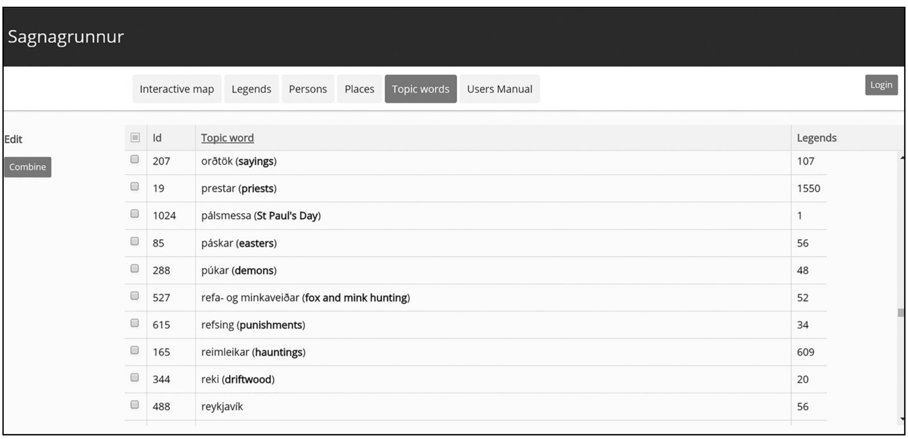

Figure 10. A quantified overview of keyword sets in Sagragrunnur. http://sagnagrunnur.com/grunnur/tags.php?lang=en (accessed September 28, 2017)

is what might not be included in the database. As a counterexample, the Dutch folktale database is a unique and refreshing exception, which quantifies its corpus of more than 40,000 texts through a variety of criteria in a clear, user-friendly way. In the right sidebar of the interface, one has metrics regarding the entire and constantly expanding stock, essentially all the major metadata in the database, such as source type, genre, subject, collector, collection time and location, names in the texts; by clicking on them, one may immediately navigate to the desired texts. Sorting is done in a descending order, starting

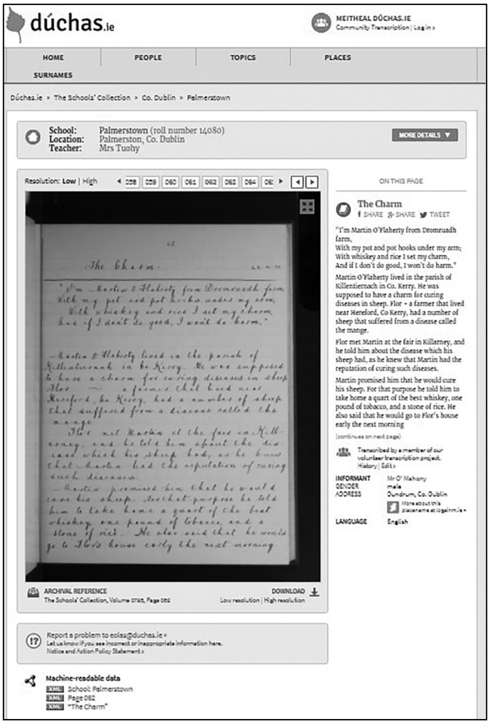

Figure 11. User interface of The Schools' Collection. https://www.duchas.ie/en/cbes/4428221/4386886/4456663 (accessed September 28, 2017)

with the most common data. For example, with regard to the source type, it is clear that texts that come from "mondeling," i.e., collection from orality, are the most numerous with 26,115 data, and among the names found in the texts, God leads with 933 hits. In the Dutch folktale database, the review of the corpus is also supported by user-driven dynamic time graphs.

Contrary to the above, Sagragrunnur provides quantified, synoptic views only in terms of keywords.

The Danish Folklore Nexus or the Portuguese legend database also do not disclose their principles of transcription, and numerous other examples could be listed. In many cases, as mentioned above, it is not even clear where exactly the text on the screen comes from. Of course, there are some positive counterexamples of this; in the Irish The Schools' Collection project, the manuscripts are scanned page by page, volunteers can also transcribe them page by page, and at the bottom of each page, one finds the exact reference to that page. They even produced a transcription guide and demo video for volunteer transcribers.[81]

In some folklore databases, e.g., Romanian love charms, Finnish and Estonian runes, ETKSpace, or the Irish The Schools' Collection, the XML download of the raw text they used has been made accessible, while others, like the Dutch folktale database, the Pan-Hispanic ballad or Portuguese legend database or the databases of the Lithuanian folklore archive do not, even though this would be crucial in terms of later usability. According to digital humanities experts, the publication of raw data and the self-reflexive documentation of the databases will be indispensable in the future, and due to the ever-changing nature of digital texts, analyses should be rerun from time to time, not published statically as a definitive outcome reflecting a certain state (Rieder – Röhle 2012:81).

The creation of scientific digital database content is a difficult, complex task that present the humanities with new challenges. At the same time, in parallel with the accurate and conscientious scientific digital textology work, there is massive digitization taking place in the world that is mainly motivated by commerce such as GoogleBooks and library science and of which it is also important to say a few words, since much of the digital content available on the web is precisely due to these efforts. The elusive, philologically unreliable and constantly changing corpus of big data circulating in the

[81] https://www.duchas.ie/en/info/meitheal (accessed August 16, 2017).

virtual space is in fact contrary to the former humanities practice where the first step in the research was to produce the most accurate and authentic text. Should we deal with these texts, and if so, how can they be used? Opinions in this regard are also divided. Jerome McGann draws attention to the disadvantages of commercial digitization and the monopolization of the field, such as GoogleBooks, highlighting the importance of producing reliable digital data (McGann 2014:20–40). According to Jockers and Underwood, however, the digital paradigm can often only reach the level of "good enough" texts, while some new types of research do not require extremely precise text editions, so working with "dirty" texts is also possible (Jockers – Underwood 2016:299–301; Price 2016). For Leonard and Tangherlini's analysis of the corpus of 19th-century Danish novels found on GoogleBooks, this kind of text quality was quite sufficient. For the texts on ETKSpace, the most important themes were identified with a subcorpus topic modeling algorithm, and then this algorithm was used to filter the corpus of GoogleBoooks for 19th-century novels that deal with "rurality" and a "rural" lifestyle. This, in turn, opened the opportunity to create a pre-selected subcorpus that would be relevant for further research (Leonard – Tangherlini 2013).

DIGITAL METHODS AND ANALYSES

In the abundance of material that digitization brought, new methods are needed to analyze the data, to see the relationships and dynamics between them. Digital humanities use a variety of methods and tools, many of which are just a computer version of an earlier practice, such as statistical methods.[82] According to Bernhard Rieder and Theo Röhle's critical survey of digital methods, the only methods that can be connected to the computer as a new medium are *stimulation, data exploration*, and *automated visualization* (Rieder – Röhle 2012:69–70). Of these methods, it was mostly the various text mining, network theory and visualization methods and tools that have introduced novelties in the digital humanities and thus in the field of computational folkloristics, which is why the examples below are introduced.

Text mining and network theory methods

The basic tools of computer analysis include various methods of *text mining.* As long as a machine-readable text is produced, there are countless options available for analysis, from simple statistics of word frequency through keywords to more complex topic modeling and latent semantic indexing (Jockers – Underwood 2016). Due to the abundance of digital data, often even the simplest, purely statistical surveys and analyses are able to put texts and corpora in a new light. However, text mining methods are capable of much more, a good example of which in the field of folkloristics is the Meertens Institute's MOMFER (Meertens Online Motif FindER), a digitized version of the Thompson *Motif-Index of Folk Literature*. The same motif index has had several online versions already,

[82] For digital humanities methods and their development, see the most recent overview: Hughes et al. 2016.

but it was MOMFER that first utilized the additional potential provided by digital methods. Apart from its multiple search options and speed, which is not just for words but also word connections, MOMFER's novelty is that the index has been expanded with a semantic search option. An English semantic dictionary, WordNet, is assigned to the lemmatized motifs of the Thompson index, which indicates syntactic relationships between words with a synonym set. Expansion thus provides an opportunity to explore more general categories and to see the relationships between each motif. According to the developers' own examples, if one searches for animals of color, one will find the green horse, even though neither green nor horse has been given as a search keyword; or if searching for poisoned fruit, the search engine will list the poisoned apple. In view of the current scientific assessment of the motif index, the creators themselves have called into question the true utility of MOMFER; nonetheless, it does provide an exemplar of an innovative, digital edition of a former folkloristic work that can be produced simply and quickly (Karsdorp et al. 2015).[83]

Text mining methods can be a major step forward in the seemingly never-ending cataloging and organizing of folklore archives and corpora. Manual annotation entails many errors that can be eliminated by various supervised and unsupervised automated techniques; moreover, the volume of the content also necessitates the indexing of content by programs (Muiser et al. 2012; Broadwell et al. 2014). As mentioned above, in the campaign of the Latvian folklore archive, one can add keywords to the typed material, which is an interesting "folksonomy" experiment, actually to find out what a folklore text collected in the early 20th century conjures up in the minds of the contributing lay volunteers, but it is quite questionable whether the Latvian folklore archive will truly be more organized through this method. Subjective, individual keywording can be a concern even in the case of archive staff (Granasztói 2008), but the new possibilities of keyword generation also gave rise to very productive initiatives. A joint initiative of the Digital Repository of Ireland (DRI) and the National Library of Ireland (NLI) has created the *MoTIF Pilot Thesaurus of Irish Folklore*, an experimental digital thesaurus of Irish folklore, which has since been utilized by the staff of University College Dublin for the thematic organization of the material of the aforementioned The Schools' Collection. In order to create the digital thesaurus, seminal handbooks, journals on Irish ethnography and folklore and works of international folkloristics have been used, from which they obtained and created a concept list through controlled retrieval (Ryan 2014a, 2014b).[84] As long as the entire content is machine-readable (in the Irish case: readable), programs can automatically associate texts with different categories.

With more complicated methods of text mining, additional novel results can be achieved. According to Tangherlini and Leonard, the analysis of the previously mentioned GoogleBooks corpus of 19th-century Danish novels is so very progressive because when topic modeling the texts of ETKSpace, the program was not designed to capture identical or similar texts, but rather "semantic similarities across a corpus based on word co-occurrence". That is why topic modeling is good at revealing topics that share the same semantic "feel" (Leonard – Tangherlini 2013:741). Text mining methods complemented by a variety of graph theory and network theory methods can

[83] http://www.momfer.ml/. (accessed April 10, 2017).
[84] http://apps.dri.ie/motif/vocab/preamble.php (accessed June 8, 2017).

be useful where the keyword method would not be able to find the similar texts. In his idiosyncratic typology, Kristensen himself classified the ghost stories of the headless horses into the category of "unnamed manor lords," but with their computer analysis, Tangherlini and his associates succeeded in classifying the text as a ghost story, even though it did not contain the word ghost or any of its synonyms (Abello et al. 2012). In another analysis of the *Danske Sagen*'s texts, they used a latent topic discovery method (LDA: Latent Dirichlet Allocation), which helped topics emerge from the corpus that did not appear in Kristensen's or Tangherlini's classification. Further research should reveal why and how these texts are associated (Tangherlini 2013a:17–19).

Text mining connects with the methods and tools of complex network research at several points. Jamshid J. Tehrani, an anthropologist at Durham University, analyzes folktales using the theories and methods of phylogenetics (study of the evolution of cognate relationships) and cladistics (methodology of tracing descendant lineages) (Tehrani 2013a). In his first study, he attempted to solve the much debated folkloristic issue of the relationship between Little Red Riding Hood (ATU 333) and The Wolf and the Seven Young Kids (ATU 123) with the help of algorithms.The article was accompanied by extraordinary interest in the natural sciences, was among the 100 most read articles in its year of publication, and was downloaded more than 73,000 times. His results were sharply criticized by the folklorists of CNRS (French National Center for Scientific Research), mainly in terms of the texts studied (58 tales, in English, thus completely ignoring the rich data of German and French variants of the type), the use of concepts (what is a motif), and the method used (NeighborNet software) (Lajoye et al. 2013).[85] Since then, he has adjusted the method of analysis and the test material in many ways to make them much more nuanced, and working with computer scientists, he attempted to reexamine another highly debated folkloristic issue. Does Little Red Riding Hood come from an oral source, that is, did Perrault (and subsequently the Grimms) adapt it from there or from a written source? Tehrani and his associates try to prove the former, and thereby refute the assertion of recent folktale research (Tehrani et al. 2015; Tehrani – d'Huy 2017). A great deal of media coverage was also generated by his joint article with Sara Graça da Silva of the University of Lisbon's Institute for the Study of Literature and Tradition, in which they attempted to determine the age of many well-known international fairy tales. For example, they claim that The Smith and the Devil (ATU 330) is six thousand years old, that is, the roots of the text can be traced back to the Bronze Age (Graça da Silva – Tehrani 2016).

Visualization

In addition to text mining methods, the most popular ones are the various *visualization procedures*,[86] so much so that, according to some, the digital paradigm is seeing a geographic and visual revolution (Presner – Shepard 2016). Of the visualization tools, it is the visualization of data on a digital map that has to be highlighted foremost, as it has significantly spread since the advent of the geographic information system (GIS). The

[85] For Tehrani's response to the critique, see Tehrani 2013b.
[86] For an overview, see: Sinclair – Rockwell 2016.

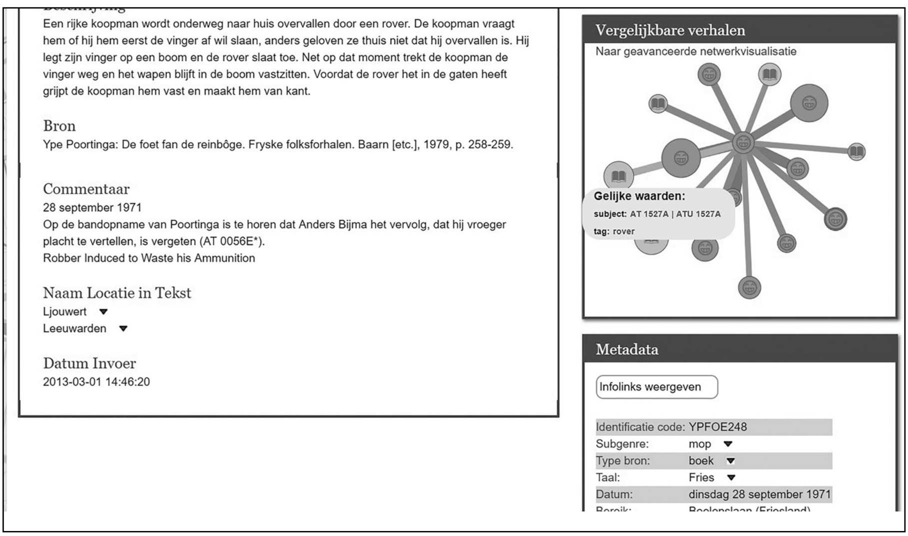

Figure 12. Dynamic visualization of the data connections of a text in the Dutch folktale database in the right sidebar. http://www.verhalenbank.nl/items/show/14643 (accessed September 28, 2017)

Icelandic folk legend database or the Dutch folktale database also use dynamic maps linked to Googlemaps. Digital cartography has widely increased the ways and means of data visualization. In ETKSpace, Kristensen's material was assigned to a 19th-century historical map from roughly the time of collection, and the frequency of a topic is displayed on heat maps (Tangherlini 2013:22–23; Tangherlini – Broadwell 2016). One can zoom in on heatmaps, where each vertical column indicates the number of texts collected in a given town: the taller the column, the higher the number. Naturally, digital visualization is not exhausted in maps. It is typically used to represent multimodal and multidimensional interfaces between data, graphs and networks. For example, in the Dutch folktale database, when one sees a specific text, one can immediately see in the right sidebar the possible connection points of the text, which can be opened right away with a click. In this case, the user can once again zoom in on the data visualization, where the location of the connection points (the farther they are from each other, the fewer common properties of the two data there are) and the thickness of the connecting lines (the thicker they are, the more matching metadata there are) can provide opportunities for further interpretation. A similar visualization solution is offered by the Icelandic Sagragrunnur for the keyword sets of legends. WossiDia also quantifies the connection points associated with individual data and displays them on the right side of the interface with a visualization that imitates the drawer system of the physical archive.

Due to its nature, data visualization is the most conspicuous element of innovation in the digital humanities. At the same time, many are pointing out that colorful, eye-catching images should be used with sufficient criticism because they can be misleading and, in contrast to an argumentation, hard to refute; a new visualization is actually needed, behind which there might be another algorithm (Rieder – Röhle 2012:73–75). Not only do some in the humanities find the digital humanities strange and suspicious because the new text editing practice entails thousands of challenges and unpredictable

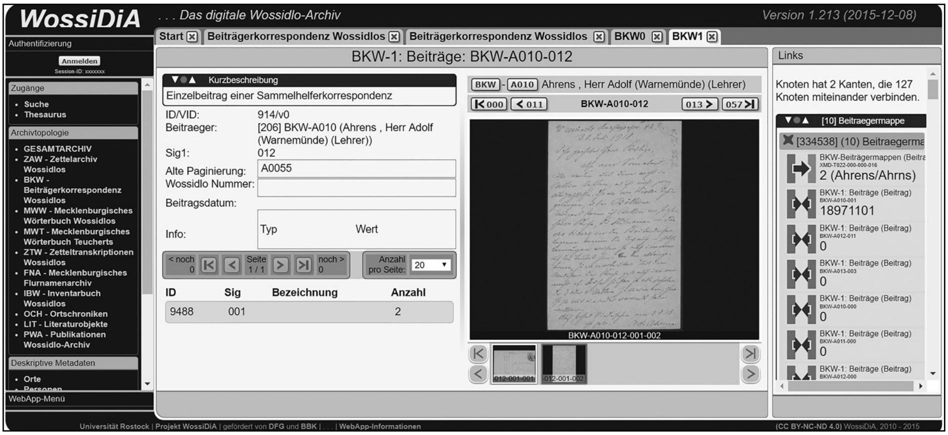

Figure 13. Data connections in WossiDIA. https://apps.wossidia.de/webapp/run (accessed September 28, 2017)

problems, but also because they have had little access to the analytical processes. The complicated mathematical calculations of algorithms and graphs may be understood by few in the humanities as they have not been trained for this. In an analysis, they only see the input and the output, not the actual analysis procedures performed by the program. Since they cannot criticize the method or the analysis itself, they are usually forced to do so with other elements of the research, such as the quality of the examined corpus and the theoretical principles of the research. Tehrani's and his colleagues' claims, for example, cannot be refuted by a traditional humanities argumentation; they can only be confirmed or repudiated by an analysis of another corpus or an extended corpus, or of the same corpus but with other algorithms and other software programs. In the absence of this, one can only point out that the motif definition they use is too subjective or that the analysis does not take into account the translation theory issues of texts translated from different languages, and so on. The phenomenon of "black boxes" is a common concern within digital humanities, and theorists of the discipline are proposing more collaborative, inter- and multidisciplinary research where those in the humanities and computer scientists work together while maintaining an ongoing discourse and genuine dialogue (Rieder – Röhle 2012:75–76; Lin 2012).[87]

THEORETICAL FRAMEWORKS

A discussion of the broader theoretical frameworks and scientific hypotheses of digital content and methods is far beyond the scope of this study, but it is necessary to touch on the subject briefly, since digital data and databases and the digital methods used

[87] With regard to the phenomenon, Paßmann and Boersma recently said that in doing research, there will always be "black boxes" that cannot and should not be opened. It is more important for a researcher to develop a competence that, although s/he may not fully understand the algorithm, would allow him/her to determine when to trust a result and when to question it. Passmann – Boersma 2017: 141–142.

for their analysis are bound to theories equally heavily, if not more heavily, than their analog counterparts. This is important to point out because technological innovation may lead some, especially the more popular readership of scientific results, to a heuristic misconception that through computer-based procedures, humanities research can finally become more objective and quantifiable, and that this hard data will make it more like natural sciences, and data-driven research will eventually put an end to the world of theories and hypotheses (Rieder – Röhle 2012; Schäfer – van Es 2017:15). Few people, even among the creators of folklore databases, reflect on how pre-determined digital content and methods themselves are from a theoretical point of view. It is only in the self-reflexive, second and third wave of digital humanities that the epistemological characteristics of digitization are more emphatically recognized. Let us look at the theoretical frameworks digital humanities research fits into, what basic assumptions it relies on, to what extent it is a paradigm changer, and how computational folkloristics relates to all this.

One of the most important theoretical and methodological starting points is undoubtedly the change of scale, which the authors capture in many kinds of expressions. The general term *large-scale* data analysis, or *distant reading,* a term used mainly by literary scholars (Moretti 2007), literary *macroanalysis* (Jockers 2013), or *folklore macroscope* (Tangherlini 2014) are all terms that try to convey the same, mostly quantitative, change of perspective, the bird's-eye view or God's eye view from where one can finally "see the forest from the tree" (Abello et al. 2012:70). The premise of the view is that, in contrast to the earlier *close reading* of texts/phenomena/data, the totality of the data can help us better understand the context of each text, and hence the individual texts themselves (Jockers 2013:27). Many have already argued that this change of scale is not the result of digital humanities. Quantitative humanities have a long history; one just has to think of 19th-century positivism or the *longue durée* concept of macrostructures that emerge from data associated with the Annales school (Jockers 2013:19; Kokas 2016:407; Jockers – Underwood 2016:292). From a little further away, the traditional historic-geographic method(s) in folkloristics also started out from similar foundations, since the collection of as many variants and as many texts as possible was a prerequisite for discerning and unraveling the story and diffusion route of an individual motif or form in order to ultimately determine the origin of these texts. According to Timothy R. Tangherlini, for this reason, digital folklore databases and computer analyses led to the emergence of a "new historic-geographic method," where the end goal is not finding the origin, the *Urform,* but rather the understanding of the latent geo-semantic relatedness of texts (Tangherlini 2013:21). Why is this important in folkloristics? What added meaning can 30,000 belief legends have from a bird's-eye view that could not be accessed before through close reading? Using the *WitchHunter & TrollFinder* program, ETKSpace researchers determined from texts and their geographical locations, for example, that stories of giants were told exclusively along large former glacier beds, or that along major trade routes, a significant number of stories survived of hidden folk/mound dwellers who, according to Danish folklore, routinely rob travellers, or that tellers consistently placed elves hundreds of kilometers away, in Jutland, which in 19th-century Denmark was still an uncultivated, wild territory (Tangherlini – Broadwell 2017:144–151). With *GhostScope & TreasureX: Conceptual Geographies*, they explored the geographic direction between individual tellers and their stories, the starting point of the analysis being that each story teller is a conceptual center, and in the case of thousands of tellers, geographic locations, directions and distances should be examined from this vantage point, too. The analysis showed that in 19th-century Denmark, most women's narratives contained references only to their actual residence or maximum a neighboring settlement. In contrast, the narratives of men, who led a more mobile lifestyle, covered a range of 20-40 km and important economic sites, too (Tangherlini 2013:20). From the relationship between lifeworlds, lifestyles and the geography of stories, it has also become apparent that while informants generally talk about events close to their place of residence, among mariners, this may increase several fold along the coast (Tangherlini – Broadwell 2017:139–140). Although the above may not seem like a remarkable result at first glance, by fine-tuning the database and formulating new questions, we can get closer to understanding the way men relate to their environment through stories.

In terms of volume, large-scaleness in digital humanities is displayed not only horizontally, but also in depth, vertically. Another key concept in databases is a *thick/deep corpus/map.* From the use of the concept, it is clear that the approach is related to the Geertzian thick description, albeit considerably adapted in its own image (Presner et al. 2014:18–19). In fact, thickness here practically means that by linking a variety of data sets, it becomes possible to analyze the same material from different perspectives, and the emphasis is on meaning-making defined by the context. According to this view, databases are never completed, thus they are always relative. At the same time, their "thickness" lies not only in the fact that they represent more data sets simultaneously, but also in the fact that they reveal different paths to understanding. Thick maps thereby bring certain natural science disciplines, for example, the former descriptive, static geography, closer to the modes of understanding of humanities and to the pluralistic interpretation of phenomena (Presner et al. 2014:18–19; Presner – Shepard 2016).

The mass digitization of data and texts and large-scale analyses can often override existing literary or folklore canons and expand the examined corpora in unprecedented ways. At the same time, it must be recognized that, currently, digitization is often haphazard, and obviously it is not only that which is on the Internet, i.e., what has already been digitized, that exists.[88] Therefore, once databases are successfully coordinated, the results of computer analyses should be treated with due criticism. As for folkloristics, for example, there are numerous unfinished initiatives and projects on the Internet, and the databases differ widely not only in terms of quality but also in terms of the volume of folklore texts. In the Romanian love charms database, for instance, there are only 119 texts, but an even better illustration is that while the Belgian or Dutch folktale databases contain more than 40,000 texts, in the Catalan database, 6,000 of the texts are only referenced, and in the French folktale database, there were around 100 texts in 2014 (Meder 2014a). How can such disproportions be offset in computer analyses?

Researchers carrying out large-scale data analyses often point out that new, computer-based facilities are not a substitute for earlier humanities research. Close reading is not eliminated by distant reading, because both are required. In fact, the combination of the two, that is, the possibility of navigating between the two, is what can truly rejuvenate research (Abello et al. 2012; Jockers 2014).

[88] On the issue of the irregularity of corpora, cf. Jockers – Underwood 2016:301.

Large-scale analyses, thick corpora, and even network theory have been part of humanities long before the emergence of digital technology, although the new medium has been extremely helpful in their development. Even in traditional humanities, their main role was to point out structures, patterns and meanings that would not have been noticed otherwise. *Pattern (frame)* is a favorite, much-used magical terminus technicus in digital humanities, although its definition raises philosophical questions, and because of the inconsistent use of the concept, it is not at all clear what traditional and digital humanities mean by it, and often not even what it can be applied to (Rieder – Röhle 2012:70; Dixon 2012). Comparing the pattern recognition of analog research with the methods of digital humanities, however, the huge difference is that in the latter, the patterns are offered by computer programs and not recognized by the researcher.[89] While in some cases patterns may be offered through computer algorithms in places where they are completely meaningless and irrelevant (Tangherlini 2013a:23), this is in fact the paradigm-changing novelty of digital humanities. At this point, we have arrived at artificial intelligence research, where the objective is to teach the program not just to read but also to “understand” what it is reading. Behind the operation of the digital methods and tools addressed in this study lies the basic research in artificial intelligence (language technology, computational linguistics, computational narratology). In practice, text mining and Natural Language Process (NLP) use a variety of methods and tools for this. Each of them needs a consistent knowledge base, such as a detailed thesaurus, a lexicon, and regular grammars. Folklore genres provide an excellent homogeneous raw material for this, a good example of which is that many of the presenters at the 2012 conference of *The Third Workshop on Computational Models of Narrative* reported on the results of folklore research, and the organizer of the conference was none other than computer scientist and artificial intelligence researcher Mark A. Finlayson, who wrote his dissertation on the machine learning of Proppian tale morphology (Finlayson 2016).

CONCLUSION

From the beginning up until the present day, folkloristics has been struggling to legitimize itself as an independent discipline. According to Lajos Katona, an important figure of Hungarian folkloristics in the early 20th century, a separate discipline is that which has its own subject and method, and the acquired knowledge forms an independent system distinct from its co-disciplines (Landgraf 2016:507). Compared with this early 20th-century notion, the inter- and multidisciplinary research of postmodernity has significantly transformed this concept of science; the boundaries of the various disciplines are today less distinct, and a great number of new, temporary forms have emerged due to a high degree of specialization. Moreover, the digital humanities seem to be radically transforming the organizational and structural way of academic life (Berry 2012; Evans – Rees 2012; Thomas 2016; McCarty 2016), while new forms of knowledge production are also emerging. We do not yet know what the text editing practice which is completely independent of book-based thinking will be like, and what the knowledge production

[89] For statistics-based computer pattern recognition procedures, see Golden 2015.

that is not thinking in books and footnotes will look like – this will only be seen by the next generation who is born into it (Debreczeni 2014:38–39).

However, the digital humanities are functioning progressively more like an independent discipline, even in the sense that Lajos Katona defined it.[90] If one looks more closely at the objectives of computational folkloristics, one can see that most of them correspond to the general objectives and areas of digital humanities (Abello et al. 2012; Tangherlini 2014, 2016). The practical aspects of digitization, the classification of texts, the issues of canons and corpora, the tools and methods used on digital content (text mining, visualization), and the theories behind them (distant reading, thick corpora, pattern recognition), are not exclusively folkloristic issues. What then is folklore in computational folkloristics? Where and what exactly is the role of the folklorist?

On the one hand, for a successful digitization, experts in the field, or to use an IT jargon, *domain experts* are indispensable. Namely, without the background knowledge and the knowledge of folkloristics and folklore archives presented in the first part of the study, it is hard to imagine the establishment and operation of a well-functioning scientific folklore database. On the other hand, a folklore database can only be designed and implemented by someone who knows not only the material but also the opportunities and pitfalls of the new technology. But this does not mean that folklorists or other humanities experts need to re-train themselves into mathematicians and computer scientists, or that in the future only folklorists that are able to write programs will be needed. What it does mean is that in the current fever of digitization, folklorists also need to be aware of the process of knowledge production and be able to utilize the new technological tools for themselves. It is important, therefore, in the 21st century that humanists be versed in content creation so that they can treat them with adequate source criticism, since the data generated and retrieved by computer programs must be used and interpreted by a researcher just as critically and rigorously as a historian would with material retrieved from an archive or other repository. The data is never passive, and data production is an active process in the digital paradigm (Rieder – Röhle 2012; Jockers 2013; Kokas 2016:412; van Es – Schäfer 2017).

In the early 2000s, the folklore database seemed to be a mere tool; in Vilmos Voigt's words, "database is a label of product, a trademark" and it could therefore not be "challenged" or even criticized because it is merely a new technological innovation "without any deeper theoretical background or profund mental activity" (Voigt 2006:308). Although the database is indeed a tool, behind it, just as behind analog text editions, lie many theoretical and methodological decisions that need to be recognized in order to interpret the results in a scientifically valid way. The focal points and patterns produced via data visualization or text mining are not the ultimate objective of research, and neither is the database itself. The syndetic, encyclopedic logic of databases, the patterns and focal points are worthless without the interpretive narrative of the researcher, since "there will always be a movement from facts to interpretation of facts" (Jockers 2013:30).[91]

In addition to acquiring adequate skills in source criticism, the main task of folklorists, knowing the opportunities afforded by computers, is to pose folkloristically relevant

[90] Major areas of digital humanities developments: 1. digital content 2. digital tools 3. digital methods. Cf. Hughes et al. 2016:151.

[91] Cf. also: Manovich 2009; Kokas 2016:412.

questions to the digital material, especially ones that would have been impossible to answer because of the earlier technology (Meder et al. 2016:93). It would be too bold to delineate what constitutes a folkloristic question today. According to Tangherlini's broad and slightly simplified definition, folkloristics has from the beginning to the present comprised the study of a diverse relationship between *people* (tellers and researchers), *places* (where they were collected and mentioned in stories), and *stories* (or folkloric expression in general) (Abello et al. 2012:65). Although there are few good examples of how to digitize the folklore archive or collection so that it will indeed become a tool for folklore research (Tangherlini – Broadwell 2014:225), the rudimentary attempts briefly summarized in this study show that researchers in the digital paradigm are concerned with and pose questions about all the folkloristic issues (textualization dilemmas, classification and typology issues, oikotypes, structuralism, diffusion and variation, intangible cultural heritage, origins, and so on) that were important in the past. With the informatization of tools and methods, individual research theories or models from the natural sciences (probability analysis and static physics, phylogenetic biology, etc.) are being reiterated. However, some authors urge us to reconsider whether adopting the theories of hard science is necessary, since humanities function fundamentally differently, and their purpose is not necessarily to provide evidence but to "develop questions and discover new insights" (Ramsay 2003:173).[92]

There is no real criticism of computational folkloristic analyses yet; its practitioners choose their words carefully and focus more on developing and testing tools and methods for the time being, as they cannot yet offer radically new questions and results. The many dilemmas and unanswered questions outlined in the study, however, were not meant to be a criticism of digital textology, and especially not a deterrent to the creation of folklore databases or computer analyses; in fact, the purpose of the study was quite the opposite. Although digitization entails many problems and pitfalls, the question remains open: how do we store, analyze and deal with folklore archives of gigantic dimensions? The digital humanities, and computational folkloristics in particular, provide a new but non-exclusive opportunity to answer these questions.

REFERENCES CITED

Abello, James – Broadwell, Peter – Tangherlini, Timothy R.

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*. Helsinki: Finnish Literature Society. (Studia Fennica Folkloristica 15).

2013 Lost in Intersemiotic Translation? The Problem of Context in Folk Narratives in the Archive. In Amundsen, Arne Bugge (ed.) *ARV. Nordic Year Book of Folklore* 69. 153–170.

[92] Referenced in van Es – Schäfer 2017:15.

Barna, Gábor (ed.)
2003 *Történeti források és jelenkori folklórszövegek lejegyzésének, átírásának és kiadásának kérdései* [The Issues of Recording, Transcribing and Publishing Historical Sources and Contemporary Folklore Texts]. Szeged: SzTE Néprajzi – Kulturális Antropológiai Tanszék.
Bárth, Dániel
2012 Historical Folkloristics in Hungary. The Past and the Future. *Etnoszkóp* (2)1:22–34.
Bauman, Richard
2012 Performance. In Bendix, Regina – Hasan-Rokem, Galit (eds.) *A Companion to Folklore,* 94–118. Malden, MA: Wiley-Blackwell. (Blackwell Companions to Anthropology).
Baycroft, Timothy – Hopkin, David M. (eds.)
2012 *Folklore and Nationalism in Europe during the Long Nineteenth Century.* Leiden: Brill. (National Cultivation of Culture 4).
Belinko, Lital – Kats, Pavel
2014 Proverbial Corpora Online. In Holger, Meyer – Schmitt, Christoph – Janssen, Stefanie – Schering, Alf-Christian (eds.) *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet,* 134–142. Münster: Waxmann. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5).
Ben-Amos, Dan
1969 Analytical Categories and Ethnic Genres. *Genre* 2(3):275–301.
1981 Introduction. In Ben-Amos, Dan (ed.) *Folklore Genres,* ix–xlv. Austin: University of Texas Press.
Berry, David M.
2011 The Computational Turn: Thinking about the Digital Humanities. *Culture Machine* 12:1–22.
2012 Introduction: Understanding the Digital Humanities. In Berry, David. M. (ed.) *Understanding Digital Humanities,* 1–20. Basingstoke: Palgrave Macmillan.
Berry, David M. (ed.)
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.* Philadelphia: University of Pennsylvania Press. (University of Pennsylvania Press Conduct and Communication Series).
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 Briody, Mícheál: *The Irish Folklore Commission 1935–1970: History, Ideology, Methodology,* 260–270, 325–331. Helsinki: Finnish Literature Society.

BROADWELL, Peter M. – MIMNO, David – TANGHERLINI, Timothy R.
2014 The Tell-Tale Hat: Reverse Engineering a Folklore Expert. DH 2015 Conference. Lausanne, Switzerland http://dharchive.org/paper/DH2014/Paper-163.xml (accessed January 16, 2017).
DÁVIDHÁZI, Péter
2014 Preface. Exploring Paradigms and Ourselves. In DÁVIDHÁZI, Péter (ed.) *New Publication Cultures in the Humanities: Exploring the Paradigm Shift,* 9–18. Amsterdam: Amsterdam University Press.
DEBRECZENI, Attila
2014 Kritikai kiadás papíron és képernyőn [Critical Edition on Paper and Screen]. In CZIFRA, Mariann – SZILÁGYI, Márton (eds.) *Textológia – filológia – értelmezés. Klasszikus magyar irodalom,* 26–39. Debrecen: Debreceni Egyetemi Kiadó.
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 BERRY, David M. (ed.) *Understanding Digital Humanities,* 191–209. Basingstoke: Palgrave Macmillan.
DOMOKOS, Mariann
2015 A folklórgyűjtővel és a folklórszövegekkel szembeni elvárások a 19. században [Expectations of Folklore Collectors and Folklore Texts in the 19th Century]. In NEUMER, Katalin (ed.) *Médiák és váltások. Identitások és médiák* II, 30–42. Budapest: Gondolat Kiadó.
EVANS, Leighton – REES, Sian
2012 An Interpretation of Digital Humanities. In BERRY, David M. (ed.) *Understanding Digital Humanities,* 21–41. Basingstoke: Palgrave Macmillan.
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*. London: Routledge. (ASA Research Methods in Social Anthropology).
FOLEY, John Miles
1997 [1995] "Folk Literature". In GREETHAM, D.C. (ed.) *Scholarly Editing: A Guide to Research,* 600–626. New York: The Modern Language Association of America.
FORRAI, Ibolya
2000 Kéziratgyűjtemény [Manuscript Collection]. In FEJŐS, Zoltán (EIC.): *A Néprajzi Múzeum gyűjteményei,* 611–648. Budapest: Néprajzi Múzeum.

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 – Nebeker, Eric
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 Wright, James D. et al. (eds.) *International Encyclopedia of the Social & Behavioral Sciences,* Vol. 23, 411–417. Amsterdam: Elsevier. (Second edition).

Golopenţia, Sanda
1997 Mapping a Network of Semiotic Systems: The Romanian Love Charms Database. *Semiotica* 114(1–2):41–66.

Graça Da Silva, Sara – Tehrani, Jamshid J.
2016 Comparative Phylogenetic Analyses Uncover the Ancient Roots of Indo-European Folktales. *Royal Society Open Science* 3(1):150645.

Granasztói, Péter
2008 Megőrzés, hozzáférés, digitalizálás: új kihívások előtt a néprajzi archívumok [Preservation, Access, Digitization: The New Challenges of Ethnographic Archives]. *Néprajzi Értesítő* XC:125–132.
2013 Az Etnológiai Archívum mint komplex gyűjtemény fejlesztésének kérdései [The Issues of the Development of the Ethnological Archives as a Complex Collection]. *Néprajzi Értesítő* XCIV:23–30.

Gulyás, Judit
2015 A szóbeliség értéke, értelmezése és a folklorisztika önmeghatározása [The Valorizations and Interpretations of Orality and the Self-Definition of Hungarian Folkloristics]. In Neumer, Katalin (ed.) *Médiák és váltások. Identitások és médiák* II, 11–29. Budapest: Gondolat Kiadó.

Gunnell, Terry
2010 Sagnagrunnur: A New Database of Icelandic Folk Legends in Print. *Folklore: Electronic Journal of Folklore* 45:151–162.

Gunnell, Terry et al.
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, 29th June 2013.) In Amundsen, Arne Bugge (ed.) *ARV. Nordic Year Book of Folklore* 69, 171–209.

Harvilahti, Lauri
2012 Finland. In Bendix, Regina – Hasan-Rokem, Galit (eds.) *A Companion to Folklore,* 391–408. Malden, MA: Wiley-Blackwell. (Blackwell Companions to Anthropology).
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 – Saressalo, Lassi (eds.)
1978 *A Guide to Nordic Tradition Archives.* Nordic Institute of Folklore, Turku. (NIF Publications 7).

Holger, Meyer – Schering, Alf-Christian – Schmitt, Christoph
2014 WossiDiA – The Digital Wossidlo Archive. In Holger, Meyer – Schmitt, Christoph – Janssen, Stefanie – Schering, Alf-Christian (eds.) *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet,* 61–85. Münster: Waxmann. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5).

Holger, Meyer – Schmitt, Christoph – Janssen, Stefanie – Schering, Alf-Christian (eds.)
2014 *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet.* Münster: Waxmann. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5).

Honko, Lauri
1998 *Textualising the Siri Epic.* Helsinki: Suomalainen Tiedeakatemia Academia Scientiarum Fennica. (Fennica. FF Communications 264).
2000a Text as Process and Practice: The Textualization of Oral Epics. In Honko, Lauri (ed.) *The Textualization of Oral Epics,* 3–56. The Hague: Mouton.
2000b Thick Corpus, Organic Variation: An Introduction. In Honko, Lauri (ed.): *Thick Corpus, Organic Variation and Textuality in Oral Tradition*, 3–28. Helsinki: Finnish Literature Society. (Studia Fennica Folkloristica 7).
2001 The digital era is here. *FFN*. 22:1, 13.

Hughes, Lorna – Constantopoulos, Panos – Dalla, Costis
2016 Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines. In Schreibman, Susan – Siemens, Raymond George – Unsworth, John (eds.) *A New Companion to Digital Humanities,* 150–170. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

Ilyefalvi, Emese
2017 Textualization Strategies, Typological Attempts, Digital Databases: What is the Future of Comparative Charm Scholarship? *Incantatio* 6:37–77. http://www.folklore.ee/incantatio/Incantatio2017_6_Ilyefalvi.pdf (accessed July 6, 2018).

Järv, Risto
2013 Estonian Folklore Archives. *Oral Tradition* 28(2):291–298.

Järv, Risto – Sarv, Mari
2014 From Regular Archives to Digital Archives. In Holger, Meyer – Schmitt, Christoph – Janssen, Stefanie – Schering, Alf-Christian (eds.) *Corpora Ethnographica Online. Strategies to Digitize Ethnographical Collections and Their Presentation on the Internet,* 49–60. Münster: Waxmann. (Rostocker Studien zur Volkskunde und Kulturgeschichte 5).

Jockers, Matthew L.
2013 *Macroanalysis: Digital Methods and Literary History* Urbana – Chicago – Springfield: University of Illinois Press. (Topics in the Digital Humanities).

Jockers, Matthew L. – Underwood, Ted
2016 Text Mining and the Humanities. In Schreibman, Susan – Siemens, Raymond George – Unsworth, John (eds.) *A New Companion to Digital Humanities,* 291–306. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

Karsdorp, Folgert – van der Meulen, Marten – Meder, Theo – van den Bosch, Antal
2015 MOMFER: A Search Engine of Thompson's Motif-Index of Folk Literature. *Folklore* 126(1):37–52.
Katajamäki, Sakari – Lukin, Karina
2013 Textual Trails from Oral to Written Sources: An Introduction. *RMN Newsletter* No.7. *Special issue: Limited Sources, Boundless Possibilities, Textual Scholarship and the Challenges of Oral and Written Texts,* 8–17.
Kenna, Ralph – MacCarron, Máirín – MacCarron, Pádraig (eds.)
2017 *Maths Meets Myths: Quantitative Approaches to Ancient Narratives.* Springer International Publishing. (Understanding Complex Systems).
Keszeg, Vilmos
2011 *Alfabetizáció, írásszokások, populáris írásbeliség* [Literacy, Writing Habits, Popular Literacy]. Kolozsvár: Kriza János Néprajzi Társaság – BBTE Magyar Néprajz és Antropológia Tanszék. (Néprajzi Egyetemi Jegyzetek 3).
Kikas, Katre
2014 Folklore Collecting as Vernacular Literacy: Establishing a Social Position for Writing in the 1890s Estonia. In Edlund, A. C. – Haugen, S. – Edlund, L. E. (eds.) *Vernacular Literacies. Past, Present and Future,* 221−235. Umeå: Umeå University.
Kõiva, Mare
2003 Folkloristics Online. The Estonian Experience. *Folklore: Electronic Journal of Folklore* 25:7–34.
Kokas, Károly
2016 Digitális bölcsészet 2016. A bölcsészek és az informatikai megközelítés: régen és most [Digital Humanities 2016. Humanists and the IT Approach: Then and Now]. In Nyerges, Judit – Verók, Attila – Zvara, Edina (eds.) *MONOKgraphia. Tanulmányok Monok István 60. születésnapjára,* 405–412. Budapest: Kossuth Kiadó.
Kolovos, Andy
2004 Contextualizing the Archives. *Folklore Forum* 35(112):18–28.
2010 *Archiving Culture: American Folklore Archives in Theory and Practice.* (Ph.D. dissertation) Department of Folklore and Ethnomusicology, Indiana University: Bloomington.
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 Logan, William – Nic Craith, Máiréad – Kockel, Ullrich (eds.) *A Companion to Heritage Studies,* 41–54. Hoboken, NJ.: John Wiley & Sons, Inc. (Blackwell Companions to Anthropology).
Labádi, Gergely
2014 A filológiai tudás formái [The Forms of Philological Knowledge]. In Czifra, Mariann – Szilágyi, Márton (eds.) *Textológia – filológia – értelmezés. Klasszikus magyar irodalom,* 173–190. Debrecen: Debreceni Egyetemi Kiadó.

LAJOYE, Patrice – D'HUY, Julien – LE QUELLEC, Jean-Loïc
2013 *Comments on Tehrani (2013).* 2013. 12. 04. http://nouvellemythologiecomparee.hautetfort.com/archive/2013/12/04/patrice-lajoye-julien-d-huy-and-jean-loic-le-quellec-comment-5237721.html. (accessed January 16, 2017).

LANDGRAF, Ildikó
2006 Archívumon innen, katalóguson túl. Többletek és hiányok a mai magyar történeti mondakutatás műfajelméleti és rendszerezési kérdéseiben [From Archives to Catalogs. Surpluses and Shortages in the Genre Theory and Organizational Issues of Contemporary Hungarian Historical Legend Research]. In HOPPÁL, Mihály – VARGYAS, Gábor (eds.) *Ethno-lore* XXIII, 27–40.
2016 A 24. óra szorításában. A folklórgyűjtés alapelvei Katona Lajos munkásságában és a korabeli magyar folklorisztikában [In the 11th Hour. The Principles of Collecting Folklore in the Work of Lajos Katona and Contemporary Hungarian Folkloristics]. *Ethnographia* 127(4):503–519.

LAUHAKANGAS, Outi
2013 The Matti Kuusi International Database of Proverbs. *Oral Tradition* 28(2):217–222.

LIN, Yu-Wei
2012 Transdisciplinarity and Digital Humanities: Lessons Learned from Developing Text-Mining Tools for Textual Analysis. In BERRY, David M. (ed.) *Understanding Digital Humanities,* 295–314. Basingstoke: Palgrave Macmillan.

MAHLAMÄKI, Tiina
2001 Some guidelines for the archiving of qualitative research data in the digital era. In *FFN.* 22(2–5):13.

MANOVICH, Lev
2009 [2001] Az adatbázis mint szimbolikus forma [The Database as a Symbolic Form]. *Apertúra.* 5(1). http://uj.apertura.hu/2009/osz/manovich/. (accessed August 23, 2017).

MARKOFF, John
2015 Archival Methods. In WRIGHT, James D. et al. (eds.) *International Encyclopedia of the Social & Behavioral Sciences,* Vol. 1, 909–915. Amsterdam: Elsevier. (Second edition).

MCCARTY, Willard
2016 Becoming Interdisciplinary. In SCHREIBMAN, Susan – SIEMENS, Raymond George – UNSWORTH, John (eds.) *A New Companion to Digital Humanities,* 69–83. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

MCGANN, Jerome J.
2010 Electronic Archives and Critical Editing. *Literature Compass* 7(2) (February 1, 2010):37–42. doi:10.1111/j.1741-4113.2009.00674.x.
2014 *A New Republic of Letters: Memory and Scholarship in the Age of Digital Reproduction.* Cambridge, Massachusetts: Harvard University Press.

2016 Marking Texts of Many Dimensions. In Schreibman, Susan – Siemens, Raymond George – Unsworth, John (eds.) *A New Companion to Digital Humanities,* 358–376. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

Meder, Theo

2010 From a Dutch Folktale Database towards an International Folktale Database. *Fabula* 51(1–2):6–22.

2014a The Folktale Database as a Digital Heritage Archive and as a Research Instrument. In Holger, Meyer – Schmitt, Christoph – Janssen, Stefanie – Schering, Alf-Christian (eds.) *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet,* 119–129. Münster: Waxmann. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5).

2014b Committee for «Folktales and the Internet». http://www.isfnr.org/files/CommitteeInternet.pdf (accessed January 18, 2017).

Meder, Theo – Karsdorp, Folgert – Nguyen, Dong – Theune, Mariët – Trieschnigg, Dolf – Muiser, Iwe Everhardus Christiaan

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 Kuismin, Anna – Discroll, M. J. (eds.) *White Field, Black Seeds. Nordic Literacy Practices in the Long Nineteenth Century,* 146–157. Helsinki: Finnish Literature Society. (Studia Fennica Litteraria 7).

Moretti, Franco

2000 'Conjectures on World Literature', retrieved 20 October 2010. https://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature.

2007 *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.

Muiser, Iwe Everhardus Christiaan – Theune, Mariët – Meder, Theo

2012 Cleaning up and Standardizing a Folktale Corpus for Humanities Research. In Mambrini, Francesco – Passarotti, Marco – Sporleder, Caroline (eds.) *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities* (ACRH2), 63–74. Lisboa: Edições Colibri.

Munk, Anders Kristian – Jensen, Torben Elgaard

2014 Revisiting the Histories of Mapping. Is there a Future for a Cartographic Ethnology? *Ethnologia Europaea* 44(2):31–47. Special issue: European Ethnology Revisited.

Nic Craith, Máiréad

2008 From National to Transnational. A Discipline en Route to Europe. In Nic Craith, Máiréad – Kockel, Ullrich – Johler, Reinhard (eds.) *Everyday Culture in Europe: Approaches and Methodologies. Progress in European Ethnology,* 1–17. Aldershot: Ashgate.

NIF

Nordic Institute of Folklore. Newsletter 1978:6(1); 1982:10(4); 1989:17(4); 1995:(23)4.

NILES, John D.

2013a Orality. In FRAISTAT, Neil – FLANDERS, Julia (eds.) *The Cambridge Companion to Textual Scholarship,* 215–223. Cambridge: Cambridge University Press. (Cambridge Companions to Literature).

2013b From Word to Print – and Beyond. *Western Folklore.* 72(3–4): 229–251.

ÓGÁIN, Ríonach uí

2013 Cnuasach Bhéaloideas Éireann: The National Folklore Collection, University College Dublin. *Oral Tradition* 28(2):317–324.

PASSMANN, Johannes – BOERSMA, Asher

2017 Unknowing Algorithms on Transparency of Unopenable Black Boxes. In SCHÄFER, Mirko Tobias – VAN ES, Karin (eds.) *The Datafied Society. Studying Culture through Data,* 139–146. Amsterdam: Amsterdam University Press.

PIERAZZO, Elena

2016 Textual Scholarship and Text Encoding. In SCHREIBMAN, Susan – SIEMENS, Raymond George – UNSWORTH, John (eds.) *A New Companion to Digital Humanities,* 307–321. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

PRESNER, Todd – SHEPARD, David

2016 Mapping the Geospatial Turn. In SCHREIBMAN, Susan – SIEMENS, Raymond George – UNSWORTH, John (eds.) *A New Companion to Digital Humanities,* 201–212. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

PRESNER, Todd – SHEPARD, David – KAWANO, Yoh

2014 *Hypercities Thick Mapping in the Digital Humanities.* Cambridge, Mass: Harvard University Press.

PRICE, Kenneth M.

2016 Social Scholarly Editing. In SCHREIBMAN, Susan – SIEMENS, Raymond George – UNSWORTH, John (eds.) *A New Companion to Digital Humanities,* 137–149. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

RAJAMÄKI, Maria

1989 Introducing Collcard. *NIF Newsletter* 17(4):35–39.

RIEDER, Bernhard – RÖHLE, Theo

2012 Digital Methods: Five Challenges. In BERRY, David M. (ed.) *Understanding Digital Humanities,* 67–84. Basingstoke: Palgrave Macmillan.

RIESCH, Hauke

2015 Citizen Science. In WRIGHT, James D. et al. (eds.) *International Encyclopedia of the Social & Behavioral Sciences,* 631–636. Amsterdam: Elsevier. (Second edition. Vol. 3).

ROGAN, Bjarne

2012 The Institutionalization of Folklore. In BENDIX, Regina – HASAN-ROKEM, Galit (eds.) *A Companion to Folklore,* 598–630. Malden, MA: Wiley-Blackwell. (Blackwell Companions to Anthropology).

2014 Popular Culture and International Cooperation in the 1930s. CIAP and the League of Nations. In HERREN, Madeleine (ed.) *Networking the International System. Transcultural Research* – Heidelberg Studies on Asia and Europe in a Global Context, 175–185. Springer International Publishing.

ROSENSTOCK, Bruce – BISTUÉ, Belén
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.
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 for the Folklore Fellows* 22:6–9.

SCHÄFER, Mirko Tobias – VAN ES, Karin (eds.)
2017 *The Datafied Society. Studying Culture through Data.* Amsterdam: Amsterdam University Press.

SCHÄFER, Mirko Tobias – VAN ES, Karin
2017 Introduction. New Brave World. In SCHÄFER, Mirko Tobias – VAN ES, Karin (eds.) *The Datafied Society. Studying Culture through Data,* 13–22. Amsterdam: Amsterdam University Press.

SCHMITT, Christoph (ed.)
2005 V*olkskundliche Großprojekte: Ihre Geschichte und Zukunft*. Münster: Waxmann.

SCHREIBMAN, Susan – SIEMENS, Raymond George – UNSWORTH, John (eds.)
2016 *A New Companion to Digital Humanities.* Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

SEIFERT, Manfred – KELLER, Hendrik
2014 Adolf Spamer online Vorüberlegungen zu einem Projekt der volkskundlichkulturwissenschaftlichen Bestandssicherung und Öffentlichkeitsarbeit. In HOLGER, Meyer – SCHMITT, Christoph – JANSSEN, Stefanie – SCHERING, Alf-Christian (eds.) *Corpora ethnographica online: Strategien der Digitalisierung kultureller Archive und ihrer Präsentation im Internet,* 85–100. Münster: Waxmann. (Rostocker Beiträge zur Volkskunde und Kulturgeschichte 5).

SEITEL, Peter
2012 Three Aspects of Oral Textuality. In BENDIX, Regina – HASAN-ROKEM, Galit (eds.) *A Companion to Folklore.* Malden, 75–93. MA: Wiley-Blackwell. (Blackwell Companions to Anthropology).

SHILLINGSBURG, Peter
2016 Reliable Social Scholarly Editing. *Digital Scholarship in the Humanities* 31(4):890–897.

SHUMAN, Amy – HASAN-ROKEM, Galit
2012 The Poetics of Folklore. In BENDIX, Regina – HASAN-ROKEM, Galit (eds.)

A Companion to Folklore, 55–74. Malden, MA: Wiley-Blackwell. (Blackwell Companions to Anthropology).

Sinclair, Stéfan – Rockwell, Geoffrey

2016 Text Analysis and Visualization: Making Meaning Count. In Schreibman, Susan – Siemens, Raymond George – Unsworth, John (eds.) *A New Companion to Digital Humanities,* 274–290. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

Skott, Fredrik

2001 "Most of your Questionnaires are Terrible to Work with". In Wolf-Knuts, Ulrika et al. (eds.) *Input & Output: The Process of Fieldwork, Archiving and Research in Folklore,* 75–114. Turku: Nordic Network of Folklore. (NNF Publications 10).

2008 Summary. In Skott, Fredrik: *Folkets minnen: traditionsinsamling i idé och praktik 1919–1964*, 281–285. Göteborg: Institutet för språk och folkminnen i samarbete med Göteborgs universitet. (Avhandlingar från Historiska institutionen i Göteborg 53).

Szilágyi, Márton

2014 Textológia, filológia, értelmezés [Textology, Philology, Interpretation]. In Czifra, Mariann – Szilágyi, Márton (eds.) *Textológia, filológia, értelmezés: klasszikus magyar irodalom,* 15–25. Debrecen: Debreceni Egyetemi Kiadó. (Csokonai könyvtár: bibliotheca studiorum litterarium 55).

Tangherlini, Timothy R.

2013a he Folklore Macroscope. *Western Folklore* 72(1):7–27.

2013b *Danish Folktales, Legends, and Other Stories.* Seattle – Copenhagen: University of Washington Press – Museum Tusculanum Press. (New Directions in Scandinavian Studies).

2016 Big Folklore: A Special Issue on Computational Folkloristics. *Journal of American Folklore* 129(511):5–13.

Tangherlini, Timothy R. – Broadwell, Peter M.

2014 Sites of (re)Collection: Creating the Danish Folklore Nexus. *Journal of Folklore Research* 51(2):223–247.

2016 WitchHunter: Tools for the Geo-Semantic Exploration of a Danish Folklore Corpus. *Journal of American Folklore* 129(511):14–42.

2017 GhostScope: Conceptual Mapping of Supernatural Phenomena in a Large Folklore Corpus. In Kenna, Ralph – MacCarron, Máirín – MacCarron, Pádraig (eds.) *Maths Meets Myths: Quantitative Approaches to Ancient Narratives,* 131–157. Springer International Publishing. (Understanding Complex Systems).

Tangherlini, Timothy R. – Leonard, Peter

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. http://dx.doi.org/10.1371/journal.pone.0078871 (accessed January 16, 2017).

2013b *Reply to Lajoye, d'Huy and Le Quellec* (2013) 2013. 12. 11. http://nouvellemythologiecomparee.hautetfort.com/archive/2013/12/11/jamshid-j-tehrani-reply-to-lajoye-d-huy-and-le-quellec-2013-5244250.html (accessed January 16, 2017).

Tehrani, Jamshid J. – d'Huy, Julien

2017 Phylogenetics Meets Folklore: Bioinformatics Approaches to the Study of International Folktales. In Kenna, Ralph – MacCarron, Máirín – MacCarron, Pádraig (eds.) *Maths Meets Myths: Quantitative Approaches to Ancient Narratives.* 91–114. Springer International Publishing. (Understanding Complex Systems).

Tehrani, Jamshid J. – Nguyen, Quan – Roos, Teemu

2015 Oral Fairy Tale or Literary Fake? Investigating the Origins of Little Red Riding Hood Using Phylogenetic Network Analysis. *Digital Scholarship in the Humanities*: fqv016. http://dx.doi.org/10.1093/llc/fqv016 (accessed January 16, 2017).

Terras, Melissa

2016 Crowdsourcing in the Digital Humanities. In Schreibman, Susan – Siemens, Raymond George – Unsworth, John (eds.) *A New Companion to Digital Humanities,* 420–438. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

Thomas, William G. III.

2016 The Promise of the Digital Humanities and the Contested Nature of Digital Scholarship. In Schreibman, Susan – Siemens, Raymond George – Unsworth, John (eds.) *A New Companion to Digital Humanities,* 524–537. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

Valk, Ülo.

2005 Establishment of the Estonian Folklore Collections and the Concept of Authenticity. In Schmitt, Christoph (ed.) *Volkskundliche Großprojekte: Ihre Geschichte und Zukunft,* 33–38. Münster: Waxmann.

Vargha, Katalin

2016 A digitális folklorisztika felé. Egy új kulcsszó és háttere a nemzetközi kutatásban [Towards Digital Folkloristics. A New Keyword and Its Background in International Research]. *Ethnographia* 127:624–637.

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). https://kb.osu.edu/dspace/bitstream/handle/1811/46903/1/FolkloreCntr_2007conference_Vastrik7.pdf (accessed January 16, 2017).

Virtanen, Leea

1993 Is the Comparative Method Out of Date? In Chesnutt, Michael (ed.) *Telling Reality. Folklore Studies in Memory of Bengt Holbek,* 255–272. Copenhagen & Turku: Department of Folklore, University of Copenhagen.

Voigt, Vilmos
1981 Computertechnik und -analyse [Computer Technique and Analysis]. In Ranke, Kurt (ed.) *Enzyklopädie des Märchens Handwörterbuch zur historischen und vergleichenden Erzählforschung*, Vol. 3, 111–123. Berlin – New York: Walter de Gruyter.
1997 Megoldott és megoldatlan kérdések hangrögzítésünk kezdetei körül. A kép- és hangrögzítés változó módszerei a néprajzi kutatásban [Resolved and Unresolved Issues around the Beginnings of Sound Recording. Various Methods of Image and Voice Recording in Ethnographic Research]. *Néprajzi Értesítő* LXXIX:103–107.
2004 A magyar folklór textológia helyzete és új távlatai [The State of Hungarian Folklore Textology and Its New Perspectives]. *Irodalomtörténet* 85:356–366.
2006 The Theory of Database in Folk Narrative Studies. *Fabula* 47:308–318.

Voigt, Vilmos – Balogh, Lajos
1974 *A népköltési (folklór) alkotások kritikai kiadásának szabályzata* [Guidelines for the Critical Edition of Folklore Texts]. Budapest: Akadémiai Kiadó.

Warwick, Claire
2016 Building Theories or Theories of Building? A Tension at the Heart of Digital Humanities. In Schreibman, Susan – Siemens, Raymond George – Unsworth, John (eds.) *A New Companion to Digital Humanities,* 538–552. Chichester, West Sussex: Wiley Blackwell. (Revised edition. Blackwell Companions to Literature and Culture).

Wolf-Knuts, Ulrika
2000 On the History of Comparison in Folklore Studies. In Honko, Lauri (ed.) *Thick Corpus, Organic Variation and Textuality in Oral Tradition,* 254–283. Helsinki: Finnish Literature Society. (Studia Fennica Folkloristica 7).
2001 Cultural Conditions for Fieldwork and Archiving. In Wolf-Knuts, Ulrika et al. (eds.) *Input & Output: The Process of Fieldwork, Archiving and Research in Folklore,* 9–24. Turku: Nordic Network of Folklore. (NNF Publications 10).

Wynne, Martin
2012 Do we Need Annotated Corpora in the Era of the Data Deluge? (Keynote abstract.) In Mambrini, Francesco – Passarotti, Marco – Sporleder, Caroline (eds.) *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities* (ACRH2), 1–2. Lisboa: Edições Colibri.

Emese Ilyefalvi earned her Master's Degrees at Eötvös Loránd University (Budapest) as a Folklorist and Ethnographer (2013) and as an Expert in Religious Studies (2014). Presently she is a PhD Candidate at the same university. In her dissertation, she examines early modern Hungarian witch trials. The main focus of her research is to understand different verbal interactions, especially the use of verbal magic. From 2013 to 2018 she has worked as a junior research fellow in the "East–West" Research Project. Within the framework of this project, she published a new Hungarian charm collection from written sources in 2014: *Ráolvasások. Gyűjtemény a történeti forrásokból (1488–1850).*[*Charms. Collection from the Historical Sources (1488–1850)*]. Currently she is working on the English edition of the Hungarian charm collection together with Éva Pócs. She also engaged with digital humanities and computational folkloristics. In 2015, she attended the Folklore Fellows' Summer School ("Doing Folkloristics in the Digital Age") and started to make an online digital database for Hungarian charms and incantations. She published several articles in Hungarian and international journals *(Ethnographia, Replika, Incantatio)* on these topics. In 2017, she was a visiting scholar for 3 months in Vienna (Collegium Hungaricum Wien), and in 2018 for 5 months in Amsterdam (University of Amsterdam). Since 2013, she has been giving lectures and seminars related to her research at Eötvös Loránd University and the University of Pécs.
E-mail: mseilyefalvi@gmail.com