doi:10.7592/FEJF2004.27.koivavesik

LEPP: THE DATABASE AND PORTAL OF SOUTH-ESTONIAN FOLKLORE

Mare Kõiva & Liisa Vesik

Abstract
The article overviews the national database of South-Estonian folklore, the online version entitled LEPP (http//www.folklore.ee/lepp), its structure, content copyrights, access, metadata and typological issues. LEPP is a universal database, containing all folklore genres and cultural history (mainly oral history). The database is free for scientific and non-commercial use. A closer look is taken at possible citation and long-term maintenance. The behaviour of large data masses and validity of search results is tested on the preliminary corpus (ca 100 000 pages of manuscript texts) compiled of the material collected from Rõuge parish.
Keywords: database, folklore, oral history, citation, preservation, South-Estonia, metadata.

The common goals of database compilers are preservation of data, conservation of source materials, systematisation of the material and making it accessible for different target audiences. *The database of South-Estonian folklore* in the form of the portal of South-Estonian folklore *LEPP. Lõuna-Eesti pärimuse portaal*, available at http://www.folklore.ee/lepp, as one of the projects of the state programme *South-Estonian Language and Culture*, follows the same principles.

LEPP is intended as a universal database, combining all folklore genres and, optimally, comprehensive archive entries, which in addition to folkloric material include cultural historical material by regions. The texts to be digitalised are systematised according to topography of origin, and represent the spiritual culture and oral narrative history from former parishes in South Estonia. As the database combines different genres of folklore, the material corpus inevitably includes items of varying format, of different scope and conditions of preservation. The data is visualised in portal format, which enables to present information of varying scope and design in form and contents, incorporate electronic publications into full data, present anthologies and selections of texts, but also develop feedback and comments applications, and enable further development.

Preparation and digitalisation of the material began with the material of the former Võru region, though selected texts from the Setu, Tartu, Viljandi and southern Pärnu region will be available for users.

To test the behaviour of larger data corpora and the validity of searches the material of one South-Estonian parish (Rõuge) was digitalised as comprehensively as possible in 2003 and 2004. A preliminary test corpus was necessary for detecting the more complex and critical problems connected with data in different formats before continuing with more extensive text corpora. A test corpus also serves other, by no means less important purposes, such as testing the system of dynamic links, determining the areas of sparse distribution of material, specifying data or gleaning additional information, but also specifying the target audience and monitoring user groups. Unlike other databases compiled at the Department of Folkloristics of the Estonian Literary Museum, which incorporate material from various Estonian archives and from the media, the first stage of LEPP is based on data from two archives – the Estonian Folklore Archives and the Estonian Cultural History Archives. Our goal in the more distant future will be incorporating, or linking to LEPP archive materials from other central archives, such as the Dialectal Archives of the Estonian Language Society of the Estonian Language Institute, the Estonian Language Archives of the University of Tartu and a part of the materials held in the Estonian National Museum. These archives contain material very similar to what is held in the archives of the Literary Museum, and is often recorded from the same informants. Without going further into the interesting ideological and strategic issues of collection methods, the importance of accumulating material from originally different archives should be pointed out in order to have a better overview of individual topics. There are very many reasons for compiling a joint database and many examples of its user-friendliness, though its materialisation requires considerable financial resources.

The article aims to provide an overview of the work carried out in the course of the project, the applied solutions and prospective future applications, the problems, problem-solving and complications connected with the database of South-Estonian folklore. The technical solutions, access to data, citations, storing the data, user ac-

cess will be examined; at lesser extent also the relations between the main data and anthological selections will be explored.

Selection of texts

Compiling a database consists of various processes, only one of which is the structural planning of the contents, writing and applying. The most time-consuming stage is selecting and digitalising the material, comparing the digitalised entries with original manuscripts, editing and systematising, specifying and editing the metadata. Another time-consuming and demanding stage is systematising the various levels of texts, taxonomy issues, information about the printed sources of texts, and other comments – in other words, the preparation of texts proper. This is possible only when metadata is presented in a standardised manner and identifiers/labels are previously agreed upon.

By the time the application for the state program of South-Estonian language and culture was submitted, preparatory work with folklore material from South Estonia had developed in various directions. Firstly, we already had previous experience in digitalising legend texts and belief reports, and knew how to work with such text materials. Secondly, we already had previous experience in compiling a general database of various genres of folklore. Thirdly, our research group was formed of expert scholars of South-Estonian folklore, which is probably the most important prerequisite for success. This fact ensured proper research approaches, valid comments and other academic additions. Fourthly, we already had studied various standards and formulated metadata.

The main tasks of the database were to support research with the entire material included in the database (in terms of parish, genre, person, typology or other taxonomy), to render available information that is otherwise overshadowed by other genres (settlement history, legal ethnography, social relations, non-traditional genres of folklore); to enable users (local schools, local promoters of culture, institutions) compile text corpora corresponding to their needs; to work with combined material; to facilitate search through various pre-selected strings (e.g. weather forecast, nix narratives), etc.

The compilation of LEPP relied on earlier experience, strategy and technical solutions of a general database, and the use of digitalised

legend and religious texts as a source material. What needed to be done was to determine the possible resources of database material, the role of additional comments, make a selection among the material that would be accessible in the database, and decide upon the access solutions. The original selection was made from among fully systematised digital folklore material and a larger amount of digitalised non-systematised material, which enabled us to draw up further strategies. Namely, since 1996 a comprehensive text corpus about specific mythological creatures and religious phenomena, which presumably included texts from South Estonia, had been compiled. These and other texts prepared in the course of digitalisation projects (riddles, phraseology) were to form the foundation of an all-inclusive database. During the first stage of the work the material of parishes in the Võru County were separated from the digitalised material. Since we lacked previous experience in working with full text databases, we decided to test the functioning of the constructed universal database on the material from single parishes and collectors (e.g. Ello Kirss) and then extend the work to the material from other parishes. Approximately 100,000 pages of manuscript material, which, owing to the network of collectors, their motivation and other factors, is relatively uneven in terms of themes and types, has been recorded from each parish. These are also the reasons why the collection of material from different parts of a parish is usually uneven.

After consulting with folklorist Kristi Salve and certain considerations we decided upon the parish of Rõuge as the test parish. Rõuge is a parish in the very core of the Võru region, with very characteristic oral heritage. Various local correspondents have worked there for years, and have contributed interesting recorded materials. The status of the model parish Karksi in the Viljandi County is similar, but also differs in that the Post-Second World War socio-political situation, however, had emptied the region of native inhabitants (Pärdi 2000) and of older folklore faster than these processes happened, for instance, on the islands; on the other hand, several archaic types of folklore, which had already disappeared in other regions by the beginning of the century, as well as certain archaic religious material, were recorded there as late as in the second half of the 20th century. The material of both of these parishes has been studied by Estonian scholars – Taive Särg, for example, has studied the folksongs and tunes of the Karksi parish. In the course of the

work came the realisation that only one test parish can be comprehensively studied, so we decided to choose Rõuge as the test parish, and the Võru County as a test region. Our original plan to diversify the database in the initial stage with the material of some colourful informant (e.g. Ello Kirss-Säärits from a settlement farm in the Setu County, who has collected unique legend and narrative lore, or Mari Sarv, a regular correspondent from the Karksi parish) unfortunately failed due to the lack of financial resources. In order to preserve the local characteristics we decided to focus on presenting the full material collected by Ello Kirss.

In the late 1990s the parallel digitalisation of several genres of folklore took place: in addition to the aforementioned legends and belief reports, the short forms of folklore were digitalised by Arvo Krikmann, Piret Voolaid and Asta Õim, also, earlier folksongs in cooperation with the Finnish Literary Society, South-Estonian Kalevala-metric folk songs within the framework of the project *Väike Kannel* by Eda-Kai Simmermann, volumes of parish anthologies of folksongs *Vana Kannel* (on Urvaste and Karula parishes), fairy tales by Risto Järv, and local lore of test parishes within the framework of the project by Heikki Valk and Mariann Remmel were digitalised. The compiled database intended to centre on religion and tradition, newer folk songs, games, non-traditional folklore genres, non-fictional lore, oral history, ethnographic accounts, biographical texts, etc. In the course of the practical work registers of persons and toponyms and other material was regularly compiled. Such progression of digitalised material indicated that mutual cooperation and reciprocal linking system enables to make more information more easily and more quickly available for the end user. This made the idea of a general database that would include more than merely folkloric material and, perhaps, less of what could be considered culture historical material (even though deciding what material it should contain proved relatively difficult) all the more appealing, and as such is easy to use for folklorists and other interested people, but also for a wider audience.

The users will be able to use a vast corpus (not genre databases), which in the broader perspective enables to observe the general issues of oral and written lore, reciprocal reflection of genres, association of certain geographical landscapes important for the per-

former, to folklore, geography of religious beings and other theoretical issues.

Language versus dialect

According to the ISO language standard, the entire material collected from South Estonia conforms to *est* or *et* standard (RFC 3066, ISO 639). Topographic definition of the folklore material would presume that the corpus would be in the South-Estonian dialect. The reality, however, is controversial: the archive material represents a diverse mix ranging from literary language to dialectal language, including different subcategories and styles of language. Variation ranges from material contributed in fiction style to oral lore written down according to the rules of written text; some entries imitate oral speech. Abbreviated versions resulting from the difference in oral speech and recording speed are quite frequent. Literary styles of the contributed materials, especially those sent in the late 19th and early 20th century, have ranged from romantic historical writings to pious literature and rudiments of critical realism. The texts also include author creation following certain structural devices and topics. Most of these texts have not been recorded in dialectal language, although there are exceptions to the rule (e.g. narratives by Jaan Sandra). In addition, reflections and generations of ethnographic elements, which are predominantly in literary language, can be found.

Next to the issues mentioned above, there are also more general problems, such as translations from one language to another. Owing to genre specifics, a part of written folklore has remained raw translations (spells, heaven's letters, practical magic, see Kõiva 1990). An altogether separate issue is the phonetic and orthographic systems created by correspondents to express the idiosyncrasies of local varieties. In case the texts happen to diverge from the correspondent's instruction, the understanding of the text is complicated, and may result in embarrassing mistakes, like the reproduction of the Setu folk songs, collected by J. Hurt, where high vocals have been repeatedly misinterpreted as North-Estonian diphthongs.

One of the principles of the database is that irregular dialectal use is retained. For example, if the dialect is absent in a recording, it will not be reproduced, because reconstruction of a living organism

is arbitrary and may yield questionable results. A reconstruction often follows a hypothetical model, and the result will also affect the linguistic data. Field work records indicate that an informant who has started speaking in literary language will return to his or her local variety, but also that in an interview, which is otherwise consistent in the use of literary or dialectal language, an informant may choose between different grammatical constructions, sometimes using them for stylistic purposes. The informant may also be characteristically inconsistent in his or her language use, introducing archaic constructions by using set phrases or quotations in his or her speech.

An altogether different issue is why the material held in the folklore archives contains other than dialectal entries. Language and dialectal language are prone to more general socio-political changes; also, their purpose, function and status have been widely different at different periods. Most of the nineteenth century recorded materials and those contributed by schools, for instance, appear largely in literary language, indicating to the status and rules of the language in the community. There are, of course, other reasons and needs that correspondents have followed or that have compelled them to behave in a certain way. Probably, one such reason is the need to preserve local intellectual heritage in literary language, by that emphasising its value, the institutionalisation of (literary) language, translation and through that interpretation of personal lore to the people who do not speak the dialect, orientation to media output, etc. An important aspect here is the general tendency of schools and organisations to acknowledge and consider literary language as standard, and cases where expert folklorists/linguists or students have failed to understand the local dialect, and are therefore met halfway and spoken (semi)literary language to, or where the informant provides the translation of what he or she said, a task not very complicated for an intelligent person. These problems have been discussed in several theoretical articles by many Estonian folklorists (Kõivupuu 2000: 202 ff, Korb 2000: 170 ff).

Dialectal languages have become consistently and increasing popular (even among people of higher education) over the last decades, this tendency is supported by the valuing of kin languages and reconstruction of the South-Estonian language variety. A dialectal language is no longer considered a combination of archaisms and inad-

equate education, but has acquired a value of its own, the ability to speak an additional language. This, in turn, has brought along a demand for accurate dialectal texts. The folklore archives and this database, however, fail to meet linguistic expectations put on them.

Editing of digitalised texts

It is characteristic of Estonian folklore archives that the majority of local contributors have been of labile literacy. Most of the nineteenth century informants and even many brilliant twentieth century informants were of poor education and inadequate literacy. Some of them had acquired primary education in a foreign language and had never been taught to write in Estonian. The twentieth century correspondents and informants include many (elementary) school children of poor competence and command of language. Owing to this, and to various other reasons, it is necessary to make orthographical revisions to the digitalised material.

Digitalisation of texts follows a standard procedure: in order to avoid recurring errors, the text is inserted, compared with the original manuscript and edited by three different persons. The digitalised text will then be revised once more, during which any remained problems will be settled.

The digitalised texts will be compared to the original manuscripts, errors will be corrected. Then follows revision according to the revising principles worked out by Arvo Krikmann and the research group of short forms (see the latest version in Krikmann et al 2001: 22 ff); these principles are analogous to the ones of earlier academic narrative editions (e.g. Laugaste 1999[1959], Loorits 2001 [1939]). In short, this means the preservation of all dialectal idiosyncrasies in recordings, though with subtle orthographical revisions: correcting obvious orthographic and punctuation mistakes (e.g. adding an upper case letter in the beginning of a sentence, adding a full stop to end a sentence, etc.). The established and collectors' personal phonetic transcriptions (mostly in the manuscripts contributed by the informants of J. Hurt in the late 19th century) will be revised in dialect form by the means of standard literary language, archaisms will be substituted with modern words, and neological forms will be evened out. Lower and upper case letters (personal names are rewritten with initial capital letters) as well as uneven stringing of

words will be corrected. The same is applied to scanned and identified typescript copies.

The transcription of audio-recordings generally follows slightly simplified transcription model of the research group of oral speech, where most stumbles, paralinguistic elements, emotions and pauses are marked, but their length has not been precisely determined, but is estimated. The same applies to video transcripts.

Database structure

The database is divided into a coherent text corpus (corpora) and metadata. The selection of **texts**, the main information included in the database, is not selective in LEPP, but should be as representative as possible – i.e. the database will incorporate folkloric, ethnographic and historical accounts, fragments of everyday life and oral history, registers of toponyms and personal names, fieldwork journals, correspondence with informants, and other manuscript data from or connected with specific Estonian parishes and included in the two previously mentioned archives (further in the course of the project the text corpus will expand with the corresponding material from other archives).

Digitalised manuscript texts will be supplemented with transcriptions of recorded materials, text-related or individual photos and drawings, musical notations, and information in other formats (sound files).

For a folklorist, folklore texts are divided into narratives, songs, religious and ritual accounts, folkloric short forms, accounts of everyday life and narrative history, various audio-visual pieces and units. We have mostly proceeded from form-centred genre labels, where narratives, for example, are further divided into folk tales, legends (sometimes even further to memorates), tales, jokes, anecdotes, etc. Each main category also has subcategories. Some categories are parallelly grouped by theme, object or subject, etc.

The problem with many shorter genres is the different background knowledge of expert scholars and common users: while in an academic publication a text is marked by a type number (and a title text), a common user searches entries by object or theme. As a compromise between the established categorisation and user-friend-

liness, the text corpus is divided into major topics, combinations of special and popular categorisation (e.g. the topic *naljad*, jokes, covers folk humour, linguistic jokes and anecdotes). A separate topic covers narratives and information which are of interest for scholars and transmitters of local culture, also for specialists, but which would render searching for the texts under typology fairly difficult (singers and storytellers, collectors, diaries, archaeology, etc.).

Therefore, the material is sorted by using the following abbreviations *j – jutt*, narrative; *vs/kk – lühivormid*, short forms; *m – muusika*, music; *uk – uskumusteade ja kombestikukirjeldus, ended*, belief report and ritual account, omens; *pilt – noodid, joonised, fotod*, notations, figures, photographs. Further possible classification will make use of standard abbreviations for a more specific folkloric categorisation, e.g. *rl – regilaul*, runo song; *ul – uuem laul*, newer folk song; *tl – tantsulaul*, dance song; *mngl – mängulaul*, game song; in terms of narratives, *mj – muinasjutt*, folk tale, Märchen, etc.(for a more detailed list see LEPP home page). The use of genre and label system should facilitate the access to the necessary information for linguists and people who might need an entire corpus on a certain topic, or more than the answer to word query.

Learning from the experience of narrative and religion researchers of earlier generations, we decided not to group mythological characters under a different genre in LEPP. Instead, texts of various length and format about denying/knowing a creature, names, tale fragments and full texts have been grouped under a provisional (legend) corpus. The texts are linked through key words. With regard to genre classification the corpus could have been divided under folk humour, legends, memorates, narratives, belief reports, expressions, etc. At the moment, the corpus of revenant, for example, includes an extensive range of revenant lore, and the plague corpus contains everything that is known about the plague, spell corpus contains various texts of verbal magic and anything related to the topic.

Example 1. Preliminary thematic categorisation:

> *Folk tales; Older tales and mythical history; Runo songs and children's songs; Newer folk songs; Round games; Singing games; Games; Music; Rhyme-book; Dances and danc-*

ing songs; Folk calendar; Weather book; Popular astronomy; Herbal wisdom and handbook of folk medicine; Spells or magic words; Wise men, witches, prophets; Dreams; Nature phonation; Jokes; Language; Riddles; Proverbs; Phrases and popular expressions; Collectors; Singers-storytellers; Nature and wild animals; Wisdom on domestic animals and husbandry; Tales from the past; Customs and rituals.

The system of key words for legend and religious texts used thus far enabled the use of up to three different key words of content per one text. As a rule, the same number of key words has been used for the LEPP database. These mainly thematic or object/subject-centred abbreviations or keywords follow the differentiation formerly used in paper files, which usually presents a combined subcategory of folklore complemented with an object- or theme-centred definition (e.g. *muis (Kalevipoeg)* – legend about Kalevipoeg; *lh (peoleo)* – imitation of the song of oriole; *uk (lepatriinu)* – religious or ritual account about ladybird). Sometimes the register includes an additional brief outline. The practical compilation of registers so far has been predominantly genre centred; the presentation of object, subject, as well as other taxonomy has been more random. With no exception, the database register presents an entry's genre and key word.

All the major categories consist of subcategories to make the database simpler for the user, providing, for example, a selection of mythological creatures and other key words connected with popular religion. Key words complementing the text enable to generate the corresponding texts from the database.

Topics, creatures, phenomena and types in specific texts are specified by standard key words and abbreviations in square brackets following the rest of the metadata.

Example 2. The key word groups together texts about the same phenomena but of different genre.

Tales from the past on various mythological creatures: Treasure-bringing goblin, revenant, devil, whirlwind, horehound, hornless ram, werewolf, Tõnn, Peko, water sprites, forest elf, ghosts, looter, incubus), *on witchcraft, treasures, apparitions, ogres (Vanapagan, Kalevipoeg, hero of Pechora, Olev), creation stories,*

horror tales, children's horror tales, stories about soul and other beliefs. Click on the key word.

Labelling works well with a specific number and determined nature of themes, creatures and phenomena, but proves more difficult when dealing with animals, birds and human activities, when it is important to determine the specific key words, or determine the precise grouping and division of the text under categories, especially in case the labelled creatures or phenomena are very specific. Would 'bird' be a suitable keyword, or is it important to specify that the entry is about a 'magpie'? Does the keyword 'medicine' needs specification that the entry is about 'wart'? Is the keyword 'spell/ medicine' sufficient, or is there a need for the same specification 'wart'?

Example 3. The entries include texts of different genres and about different phenomena. In narratives and religious accounts these are determined by means of genre and key words

vesi / ns – vetehaljas, tekstis ka loits (water sprite, in the text: also spell)

järv / pn – järvemuistend, tekstis ka kohanime saamine (lake legend, in the text: also the origin of a place name)

mägi / koduk / aj / kal – jutt mäest, tekstis kodukäijast, kalendritavandist ja kohalikust ajaloost (a tale about a hill, in the text: also revenant, calendar ritual, local history)

Following the labelling of basic categories (song, tale, beliefs, language, etc.) and key words has been based on and followed according to genre taxonomy. The most problematic aspect of this work is presenting the typology consistent with international or national type registers. Determining the type depends of the previous work on a text corpus. Typology, in principle, and particularly if it follows an internationally recognised system, enables to locate analogous texts from the lore of different countries. Unfortunately, there are only few analogous universal typologies, as most of these are limited to the lore of a single country. In Estonia, typologies have been created for folk tales, spells, proverbs, riddles, but also on phrases and popular expressions, partly also on legends, folk humour. Pille Kippar has compiled a concise type register of animal folk tales,

covering all the manuscript and printed material (Kippar 1986). Rudolf Põldmäe and Arvo Krikmann have put together the typology of older folk humour, which is also available on the Internet (Krikmann 1999). The typology of folk tales and legends follows Antti Aarne's type register of 1918, covering only Jakob Hurt's manuscript files (Aarne 1918). Newer manuscripts contain tales that are not included in Hurt's files and that have been classified according to the descriptions in international type registers (Uther 2004; Thompson 1961, Jauhianinen 1998, Simonsuuri 1961). Typescripts and register of folk songs include type names provided by Hella Wuolijoki, and their contaminations have been determined by Hilja Kokamäe, Herbert Tampere, and others on the basis of Wuolijoki's work and these are consistently used in LEPP. Newer folk songs have been grouped into larger thematic units according to the register typology by Ingrid Rüütel. The original types and groups of legends have been determined by Ellen Liiv on the basis of Lauri Simonsuuri's legend register, later also on the basis of Marjatta Jauhianen's improved edition of Simonsuuri's work. The latter typology has been included in the LEPP database only partly.

Since typology is the most important tool for a folklorist, the major question in compiling a folklore database will be whether it is necessary, or even possible, to present a typology. Typology will become evident only in consistent preparatory work, if the results are reflected in the manuscript theme register, typescript copies, type register or the preparation has been widely different. While the folk song register includes song types and merged types, the same types are listed under different type names in volumes of parish registers, which largely depend on the preferences and background of a specific compiler. A folk tale, for example, may be categorised under a (main) type, whereas its contaminations are usually not. In reality, typology is often available only in the academic publication or paper files, and not in the joint register, which has not been updated due to the lack of time. It is not an easy task to compile a typology of these folklore genres that do not have one, as it requires years of (preparatory) work. Typologies of single genres created thus far have taken, for example, a team of five folklorists 25 years at proverbs, and one scholar almost 25 years at Estonian animal folk tales. Compilation of other major corpora has also been extremely time-consuming – the digital text corpus of printed Finn-

ish Kalevala-metric folk songs, for example, has taken six scholars three years and will apparently take three more. Thus, determining and creating accurate typological units is less effectual in a mixed corpus containing different types of texts and is analogous to constructing folk song typologies of single parishes – a single type may have as many different type names as there are published folk song volumes. Thus we assume that typological issues require detailed examining and more optimal solutions. Therefore, LEPP now includes typological units only if they have been determined in advance, and the facilitating tools are mostly key words and links referring to the context. Hopefully in the future information can be found through search engines as well as through dynamic links (Elango et al 2003) and by other means.

An altogether different issue is whether a scholar of other than folkloric background will be able to learn to work with specialised typology and which cognitive or explicative function does it hold for him or her. Or is it limited to the skills of operating a video camera, which can be used automatically by everyone, requiring no knowledge of technical details and recording mechanisms?

Metadata

In terms of metadata, the requirements of databases include one-to-one function (http://dublincore.org) and standardisation. An important part of textual contents is metadata. Metadata includes information reflecting the contents of the text (key words, taxonomy and categorisation, text typology), but also identification information of the text by time, space, person, etc. In addition, the database metadata inform of the legal data (performer, collector, publisher, people involved in preparation work, project, prior publication in print, authenticity, etc.) and database-specific information (format, technical data).

Since the database includes texts of different length, function and format, the metadata is presented by means of the adapted version of the Dublin Core Metadata Element Set (http://dublincore.org/documents/dcmi-terms/#H2). Application standards of e-publications, comments and links and metadata used in the portal format meet the general requirements, including information on ISBN, ISSN,

URL, compiler, editor, year, project, author rights, financer, technical effecter and technical information.

Folklore archive texts follow systematisation principles that have been established a long time ago and have changed very little over the past one hundred years. The main change is related to the requirement of presenting as accurate data as possible, another one is the traditional identification of entries by enumeration. Neither of these changes is anything new, and have became particularly important in the late 1920s and 1930s fieldwork results by scholars and stipendiaries of the Estonian Folklore Archives. During this period the texts were furnished with as accurate topographical and personal data as possible. In the period following the World War II the standard became less strict, and the accuracy depended on specific collectors and decades. Identification numbers were dropped in this very period, and this has resulted in the uneven identification of texts.

Thirdly, research groups preparing academic text corpora and databases proceed from different principles and have expressed their views in publications and at copying campaigns. The uneven nature of data has been one of the main reasons for additional work at compiling modern databases.

Each specific text entry is given an archival code, consisting of the abbreviation of the manuscript file, format and number of the manuscript, followed by the number of page and piece. Traditionally, several manuscripts are not referred to by the entire number code, but rather by the page number (manuscript files of Mattias Johann Eisen, Samuel Sommer, Walter Anderson); there are also other ways of reference, diverging from the established norm. Since the archives tend to alter the abbreviations of archival items and principles of reference, it would have been practical to revise the abbreviations of manuscript folklore files in the initial stage of digitalisation (it is possible, though, that this can still be done).

Each entry's place of collection is specified as accurately as possible up to the farm level, by using the Place Names Database of the Estonian Language Institute (http://www.eki.ee/knab/knab.htm).

Full collectors' names, including the first name and the surname, are presented. Specification of names and pseudonyms, and evening

out the possible name variations is based on a list compiled by Rein Saukas (2001), and generalises the proverbial and riddle material thus far. The list has been revised and improved during digitalisation of legend texts and belief reports.

The names of informants are also presented in full, with the additional former name, maiden name, etc., if such information has been included in the material. Informant's date of birth or age also belongs among the required information. Occupation is included only when it has been mentioned in the collector's notes.

The year of collection, or the year of when the text was recorded and entered into the archive, also belongs among the main metadata.

Inclusion of legal metadata guarantees the rights of people involved in the creation process, such as performer, collector, digitalisers, publisher or compiler, project, foundation.

While the archival data up to the time of collection is an inherent part of the text and is generally a public information (unless the performer has specifically asked not to publish his or her name), available and searchable for every user, then the data connected with digitalisation is usually core information and becomes available through the advanced search. At the moment the digitalisation data is also available in LEPP. An important role in working with the digitalised data is played by people involved in the process and the time it was carried out, as it is their work and interpretation that will reach the end user: the name of the person who entered the information and the time it was done, the name of the editor and the time of editing, and the name of the second editor-reviser and the time of the second editing.

An important aspect is informing of principles of editing, project requirements and legal rights. LEPP's outcome will be publicly available for scientific and educational purposes, though the project's title name has not been included among the metadata as yet. With the increase in the number of similar projects, there may be a need to add the general title of the project to the digital text.

While in folklore authorship rights of performers have been strictly followed, the archives have been more negligent in terms of authorial rights of collectors and photographers. This, however, is also a

cultural property of the collector, on which he has taken on work and ethical responsibility, and sometimes even financed the outcome.

The internal and external information mentioned above concerns the additional information added to the text. Presentation of maximum metadata will become available analogously to library search systems, opening a large number of bibliographical details/ metadata of texts only as a result of advanced search.

In addition to digitalisation data, the advanced search will also enable access to information about the previous publications of the text, scholar's assessment on whether the text is authored, a copy or an authentic text, structural and logic formulae put together by scholars, etc.).

Example 4. Metadata with digitalisation information.

> *H III 11, 175 < Rõuge khk. - Jaan Gutves < ? (1889). Sisestas Eva-Kait Kärblane 2001, kontrollis Kadi Sarv 2002, parandas Salle Kajak 2002 [mägi / pn]*
>
> H III 11, 175 < Rõuge parish - Jaan Gutves < ? (1889). Entered by Eva-Kait Kärblane 2001, first revision by Kadi Sarv 2002, second revision by Salle Kajak 2002 [mägi/pn]

At the moment, information about the publication of the text in earlier sources has been included wherever possible. This information is included directly after the main archival data and before digitalisation data. Printed publications will be referred to by the name of the editor, title of the publication, heading of the text, time and place of publication, number of pages. This information, as a rule, is followed by other comments related to the contents, type descriptions, etc.

Example 5. Wherever possible, information about prior publication has been included.

> *ERA II 160, 372 / 3 (31) < Rõuge khk., Pindi v., Pindi as., Karja t. < Rõuge khk., Vastse-Nursi v. - Ija Daniel < Peeter Märdimäe, s. 1865 a. (1937). Avaldatud O. Loorits Endis-Eesti eluolu II, Tartu 1941, lk 52. Sisestas Salle Kajak 2001, kontrollis Kadi Sarv 2002, parandas Eva-Kait Kärblane 2002 [mets / bot]*

ERA II 160, 372/3 (31) < Rõuge parish., Pindi p., Karja Street < Rõuge parish., Vastse-Nursi p. - Ija Daniel < Peeter Märdimäe, born in 1865 (1937). Published in O. Loorits Endis-Eesti eluolu II, Tartu 1941, pp. 52. Added by Salle Kajak 2001, first revision by Kadi Sarv 2002, second revision by Eva-Kait Kärblane 2002 [mets/bot]

Database-related information is more laconic in texts (appear in simple text file format in the database), and more extensive in sound, video, photo files, where the information about the recorder, recording format, and device, etc. will be included.

So far, the notations, photos and figures digitalised for LEPP have been included in separate tables of the database, owing mostly to non-standard formats and problems connected with further linking. .tiff format and two different sizes have been used for pictures and figures, video clips are in mp3 audio format.

Technical description

LEPP is programmed by Saamuel Vesik by the means of MySQL, which is a flexible database, independent of commercial platforms, and compatible, enabling to combine data of different format and size. This became our choice mainly because the database complies with other databases and text corpora and enables to transfer data should a need arise. We have used PHP programming language and XHTML Web output.

Since the financing of the project was short-term and covered the work of a part-time data enterer, in some years also partly the work of a part-time reviser and some elementary programming, wecame to a compromise, in the view of the imminent result, and postponed the use of PostgreSQL. The chosen MySQL and .PHP solution ensures the calculation of the core values of the database, but enables to achieve results and display the data in the Web simpler and faster. Data included in the database can be easily edited, and are also easily convertible. Users will need to install no additional programmes or additional software to access the material. Data validity to XHTML standards is also automatically assessable and validation results will be supplied with a Web note, analogous to the e-publications linked to the portal, which HTML and XHTML-format

validity has been revised through automatic validation. A corresponding note will be added to the specific publication. The revision date of all the provided links will be added, thus ensuring reliability.

We have attempted to make all the solutions as simple as possible, enabling to read and download texts of even form: archival data in the header is followed by the text, texts are separated from each other.

On a passing note, we have also made attempts to make the database easy to use for people preparing the database and entering the data, since their work demands great accuracy and attention, but is also routine and usually time-consuming; the same text is repeatedly edited by different people.

Information included in the LEPP database follows the methods used for digitalising and downloading legend and belief reports: metadata is entered in an ordinary word processor file, LEPP converter adds the metadata and texts from .txt files into the corresponding database fields. The converter distinguishes text file information by customary markers in the archival code, for example, key words and classifiers in square brackets of the header; data on entering and publishing and comments on typology from the space between square brackets and parentheses; year(s) of recording in parentheses; an origin mark distinguishes between the informant and topographical data (*khk,*– parish, *v.* – rural municipality, *k.* – village, *t.* – farm, etc.). The dash separates the collector from the header. The text is separated from metatext by a section sign; texts are separated by two section signs. The corpora of the material from each parish are presented in different tables, owing to the considerable size of the material. Texts of two parishes of origin are for the time being included in the corpus of both parishes.

Estonian folklorists apply different methods to enter material into databases. Various projects make use of ready-made forms, which enable entering the metadata as well as database texts (*Justkui.* Database of Estonian Phrases http://www.folklore.ee/justkui, Baran et al 1998-, Baran et al 2003; Database of Local Lore using MS Access, Remmel 2002). Entering data and texts as, say, Excel sheets, or converting from MS Word into Excel tables is almost as effective,

since mistakes in filling the fields are readily revealed. This option has been used at compiling databases of riddles (Krikmann et al 2002) and of the periphery of riddles (Voolaid 2002, 2003). Texts and metadata entered in ready-made forms are more difficult to check and revise, since the previous and the following texts can be not observed at the same time, copying similar data is more time-consuming, etc. The drawbacks of online revision are partly caused by the acquired horizontal reading habit, and thus lapses in attention, which obstructs detecting and therefore also correcting errors in vertical screen text. Text and computer screen are usually located on different levels, which is why the revision of digitalised legend and belief reports is carried out according to the traditional revision method of publishers, where one person reads out the manuscript and another one revises the digitalised text).

The number of mistakes is directly related to the attentiveness of revisers while using the converter. The first input in the LEPP database revealed that ca 50 of the 6,000 texts received error message, most problems were caused by disregard of certain key words. Even though looking for and correcting errors is an extremely tedious task (and uncomfortable, due to their position), the converter facilitates the merging of revised texts to the database.

A user can read a typical text file: a header followed by the text, with different texts separated from each other by space, as indicated above. Ten or more texts are displayed on the screen at a time; the rest will open on a mouse click on the page number.

Database in portal format

Since database format preconditions a compilation of a text corpus and a query system based on that, it appeared practical, considering the alienation of contemporary users and time resources, to join the database with a portal format, enabling presentation of data of varying scope, alternation of news with longer texts, online forums and interactive communication, if necessary, open and merge galleries and music selections, e-publications of the corresponding region (the Setu folk songs, epics of Anne Vabarna, etc.) with the historical description of a parish and a pre-selected list of links.

The Web page of each particular parish contains archival texts and additional links to selected materials on the topic, or those expanding the material which is not directly connected with the entries of either of the archives, but function mostly as an aid in memorising and providing additional information. The location of a parish can be determined through the maps of parishes designed by Andres Kuperjanov (http://www.folklore.ee/rl/folkte/kk/maps/we.htm). We have also added a chapter from the 1920s-1930s parish anthology, which provides information on the landscape, history, architecture, demography, social activities, and an overview on folk culture (http://www.folklore.ee/lepp/rouge/tutv.php). Contemporary monuments, administrative information and culture are introduced in a minimal number of links:

http://www.matk.ee/tegevus/Yritused/Kiidi/rouge.htm http://hot.ee/rauge/ , http://www.rk.ee/symb/rouge.html; http://www.diaso.ee/fotod/vaated/louna_eesti/01.html (a gallery of picturesque nature photos).

Of course, beautiful photos and additional visual information is also available under the subcategories of the database, which include also links to other folklore databases (droodles, phrases, conundrums, etc.).

A separate network is formed of links to various e-publications and anthologies of regional lore. The Web page of the material on the Setu region, for example, includes links to complete volumes 1-3 of *Setukeste laulud* (The Setu Folk Songs) by Jakob Hurt and Paul Hagu (http://www.folklore.ee/rl/pubte/ee/setu/laul/), as well as the Web page of the great twentieth century Setu singer Anne Vabarna, compiled by Paul Hagu (http://www.folklore.ee/rl/pubte/ee/setu/anne/), and the epics Ale (http://www.folklore.ee/rl/pubte/ee/setu/anne/ale/index.html)) and Peko 2 (http://www.folklore.ee/rl/pubte/ee/setu/anne/peko2/index.html), composed by Anne Vabarna. Both epics are supplied with a short introduction and creation history.

There certainly rises the question, why do we need anthologies, selected texts, biographies of performers, etc. when we already have this extensive database where analogous data is available? The main advantage of the commented extracts and authorised biographies is that they have been made available; also, they enable to find sys-

tematised information more quickly and provide complete overviews instead of single facts. Each analogous subdivision or subcategory has its own value, as the subcategories represent texts systematised according to scholars of different periods, including their comments, analysis results, hypotheses. In many cases the selection provided by the search engine required additional data for further use and interpretation, or earlier treatments of the same subject. Here the constantly expanding server of folkloristics, together with its e-publications, has been of valuable help.

Preservation

The long-term lossless data preservation is the central and complicated issue in digitalising material. The main reason for this is that data carriers have changed quickly over the past few decades, and the expenses and time spent on transferring digital information from one format and data carrier to another have been considerably large (Kuperjanov 2005). To name only the most general quotidious problems: the physical endurance of the new generation hard disks has decreased contrary to their increased capacity, software does not last long and requires consistent updating to comply with the newer versions and programming languages. Since Web output is important in providing a better service for the database users (including the LEPP portal), the database compilers are inevitably faced with ever-changing standards and opportunities. Many attractive interactive and dynamic solutions are often particularly short-lived and depend on the altering browsers, fashion, etc. (Bird & Simmons 2002). Updates of the software of the server of folkloristics has indicated the inconveniences the situation has caused – sometimes it has taken server administrators weeks to ensure proper working stability.

Proceeding from the above it is imperative to (i) preserve data in as compatible format as possible, (ii) regularly update server software and observe that resources are available and function properly, (iii) create different automatic and individual backup copies on different data carriers.

The compatibility of South-Estonian texts is guaranteed by the fact that the texts are preserved as .txt format original files. As indicated above, each single entry comes with metadata and markers,

and is clearly separated from other texts by signs which the engine is sensitive to. The file system is analogous to that of paper files, i.e. materials digitalised from a specific paper volume have been digitalised in the same file, and the archival code of the paper material is included in the file name. A complete file usually contains material originating in different parishes and collected or recorded from various persons. These files are held in folder e-archive, in the internal server OHTO and for security reasons the entire e-archive is automatically copied on DAT tape each night and on reusable DVD once in every three months.

The database is located in the external server Haldjas, where all complete text entries are uploaded and e-publications are held. Each night a backup copy of the information on Haldjas is created on DAT tape; in addition, larger databases and their scripts have been separately copied on DVD. Furthermore, the entire content of the server disk is mirrored to another hard disk, so that the data can be reproduced, should one of the disks malfunction. Once a year, the system administrator is obliged to copy the entire contents of the external server on DAT tapes.

Unlike several other databases of the Department of Folkloristics, the LEPP database will be updated only a few times a year, that is, it will not be done regularly. The advantage of this approach is that it will be easy to make accurate backup copies, and the main drawback is that it would require an extremely effective file documentation to be able to find an accurate uploading sequence. An additional advantage is that there is enough time to compare the digitalised material with the original manuscript and make revisions in the meantime. The files have also been printed on paper and are preserved in folders. Mare Kalda has initiated the including of short reports on the work with the files.

Citation

LEPP is ISBN numbered, which means that the entire material with its subcategories, available on the portal Web site, is registered in the national ISBN centre. The joint server of folklorists holds a number of data which is indexed or ISBN numbered. Another reliable identifier is URL and the permanent Web site ad-

dress (http://www.folklore.ee/lepp). Movement of Web sites is strictly avoided within the server. All the articles published in electronic journals have unique Web site addresses, including .HTML or .PDF titles, which are also included in the printed versions of the journals. Such data enables access to and correct use of the articles. Data held on the server has been moved only on extreme occasions, in which case alias has been used to ensure that the user will be transferred from the old site to the new one, by that guaranteeing the general verifiability of data. File names are never changed. An appointed person will be responsible for the functioning of aliases.

Since searching, digitalisation and preparation of the archive texts for the database is a copious work, the user will be expected to include a reference to LEPP and its URL, which was used to find the text, in the list of manuscript sources and references. An author can thus use the original text and its archival reference, provided in the header of each text entry, for citation, but like printed publications, the user is expected to refer to the sources of the quoted material and not claim to have used the original archive manuscripts.

Bibliography of the LEPP database is presented as a list of folklore files and other culture historical foundations (and their abbreviations), which were used for entering the texts, also the list of digitalised publications. The figures of entering and revising the material in the database are presented in a chart which complements the bibliography and provides a visual overview of works currently in progress.

As already indicated the entries are supplied with unique archival references, or character and number combinations, which should prevent overlapping and accidental overwriting or deletion. This number combination helps to link a digitalised text to the corresponding manuscript. Should a necessity arise, the unnumbered single texts have been given a number during the digitalisation process, differentiating these from other texts of similar number combinations (this will become important if the collector and/or presenter, or both, are the same, or if a manuscript page contains several numbered text units of similar contents, such as e.g. the series of belief accounts in manuscripts by Eisen). There is a gen-

eral need for a unique archival code, but it becomes critical only if the metadata accompanying the text proves to be inadequate: for example, there is no information about the collector, or the place of collection, etc., and the entry has to be identified on the basis of the archive reference and key word.

Another important criterion is the stability of the digital source: completed e-publications and other electronic materials are preserved in the form of completion. If major changes are made to the database it has been pointed out that "it is the improved, edited and updated version", a new edition with a new sequence or registration number is registered, analogously to printed publications, and the original version is preserved in entirety. The offline version of the BERTA database (http://www.folklore.ee/berta/), for example, is now, after several major improvements, available in a new version, and the changes in the online version are pointed out on the main page. The nature and time of changes to the database are pointed out.

Rights

User rights have been specified in detail on the project title page and are analogous to the main rules governing the external Haldjas server: the user has free access to the material prepared in the course of the project, which means that the user has the right to view, download, print, use the material for educational and research purposes, distribute it for non-commercial purposes, link to the material available on the site. The lore corpus can be freely used for academic research, and is prepared for being used in academic publications; it can be used for educational purposes and for promoting local cultural life. The use of the texts for commercial purposes is strictly forbidden.

Project coordinators should be notified of using the material, Linking to the texts of the LEPP database can be notified by e-mail, it is also recommended to inform of the use of the copied material in the Internet. Special contract needs to be signed for the publication of the database material.

Established principles of quoting should be followed in research papers and various printed publications by including the title and

location of the source – both are presented on the project title page. The printable version includes the project title and Web address at the beginning of each page.

Conclusion

In current form and with updated contents, we estimate that LEPP can be used for a long time. The portal and database were expected to make available a positively large amount of information on South Estonia. The objective and subjective strengths and problems of the project have long been evident. The main concern is the struggle with short-handedness. Work at the database would require the continuous contribution of at least two or three people, with the regular additional help provided by the programmer. Currently, the project has only covered the work of a part-time enterer of data, and a reviser, other works have been done if time allows next to the regular work of the Department of Folkloristics. We gratefully acknowledge the support provided by the Estonian Cultural Endowment, and members of the Group of Folk Belief and Media of the Department of Folkloristics at the Estonian Literary Museum.

Digitalisation, text revision and editing were strongly affected by the closing of folklore manuscripts for use, after which further work was done by means of microfilm readers. Apart from the increase in the time required for completing all the tasks, there are also readability issues connected with manuscripts, health hazard, and other problems. Since there are no alternative choices to read the original texts, the databases will be extremely valuable, for they ensure access to the material in orderly form. Making the texts available will definitely increase their being used.

During compiling the LEPP database came the realisation that linking the material of different projects is far more complicated than we originally thought. There are various obstacles, like the variability of general digitalisation standards, varied formulation of metadata, comparison of digital and manuscript data, as well as uneven observance of revision principles. The latter ranges from the orthographical revision, formulated by Krikmann, to the reproductive revision through reproducing the original manuscript or modifying a dialect. The general inclination to proceed with various databases and academic publications calls for searching for the same,

or similar, but at least compatible solutions. Comparison of digitalised material with the original texts and final revision as a demanding task in terms of both time and qualified people for the work tends to be neglected due to the lack of resources. Only the most necessary tasks are correctly revised, such as the material of printed publications. This, again, addresses the general under-financing of projects, leading to the slow time of completion and quite a number of problems connected to and proceeding from this. A target financed project usually has a project manager responsible for preliminary studies, as well as part-time data enterer(s). One of the main technical problems at the present moment is preservation of data in different formats, some of which are not easily convertible or made available in the Internet.

Owing to its scope, LEPP will be "under eternal construction" for some time in the future, but the existing material is already available for use. One of the major achievements of the database and portal is the digital corpus of the Rõuge parish, a preliminary test corpus which enables to solve major issues and develop further application of the database, including making the material more consistent, which should solve the problems affecting the flexibility and speed of queries, but also finding simple solutions for enterer(s) and end users. In the nearest future, the intermediary versions of LEPP containing the material of single parishes, as well as the completed e-publications will be replicated as offline versions or on CD-ROMs. User-friendliness is also an aspect to aspire to, so that the material would be available to people unfamiliar with folklore taxonomies and typologies. One of our next steps will be monitoring user audience and adjusting technical solutions based on these results.

The LEPP database project group expresses hope that the interest of the general public in local traditions and open-mindedness of the society towards supporting major application projects will not fade, thus enabling the development and promotion of full text databases and services.

Acknowledgements

This article was written within the statefinanced topic 0032472s03 and with support from the ETF (grant no. 5117). Salle Kajak (2002-2004) and Eva-Kait Kärblane (2001) were responsible for entering data. Other people involved in the project include Mare Kõiva (project manager, labelling, selection of texts), Kaljo Põllu (design), Saamuel Vesik (database, PHP and XHTML), Kadi and Pihel Sarv, Anne Kaaber, Ell Vahtramäe (editing texts), Mare Kalda and Luule Krikmann (revision of texts).

References

Aarne Antti 1918. *Estnische Märchen- und Sagenvarianten: Verzeichnis der zu den Hurt'schen Handschriftssammlungen gehörenden Aufzeichnungen mit der Unterstützung der Finnisch-ugrischen Gesellschaft.* Folklore Fellows Communications 3: 25. Hamina: Suomalainen Tiedeakatemia.

Baran, Anneli & Hussar, Anne & Õim, Asta & Õim, Katre. Compilation of the Database of Estonian Phrases. *Folklore: An Electronic Journal of Folklore* 25, pp. 35–72 (http://www.folklore.ee/folklore/vol25 – August 10, 2004).

Baran, Anneli *et al.* 1998–. *Justkui: Fraseoloogia andmebaas* [Justkui: Database of Estonian Phrases]. Tartu & Tallinn: EKM Folkloristika Osakond (http://www.folklore.ee/justkui – 10. august 2004).

BERTA offline-versioon 2004 = Kõiva, Mare & Särg, Taive. *BERTA: Eesti rahvakalendri tähtpäevade andmebaas* [BERTA: Database of Estonian Folk Calendar Holidays] (http://www.folklore.ee/Berta/cd-versioon – August 10, 2004).

Bird, Steven & Gary Simons 2003. Seven dimensions of portability for language documentation and description. *Language* 79, pp. 557–582.

Dublin Core Metadata (http://dublincore.org/documents/dcmi-terms/#H2 – August 10, 2004)

EKI kohanimeandmebaas (KNAB) [Place Names Database KNAB of the Estonian Language Institute] (http://www.eki.ee/knab/knab.htm – August 10, 2004).

Elango, Aravind & Bollen, Johan & Nelson, Michael L. 2004: Dynamic Linking of Smart Digital Objects Based on User Navigation Patterns. *arXiv:cs*.DL/0401029: vol. 1 (28 Jan) (http://xxx.arxiv.org/PS_cache/cs/pdf/0401/0401029.pdf – August 10, 2004).

Hagu, Paul 2002. *Anne Vabarna.* Tartu: EKM USN & Tartu Ülikool. (http://www.folklore.ee/rl/pubte/ee/setu/anne/ – August 10, 2004).

Hurt, Jakob 2002. *Setukeste laulud* I–III [Setu Folk Songs]. Hagu, Paul (ed.). Tartu: EKM USN & Tartu Ülikool (http://www.folklore.ee/rl/pubte/ee/setu/laul/ – August 10, 2004).

Hussar, Anne & Krikmann, Arvo & Saukas, Rein & Voolaid, Piret 2001. *Eesti mõistatused* I: *1–1350* = *Aenigmata Estonica* I: *1–1350* [Estonian Droodles]. Monumenta Estoniae Antiquae IV: 1. Tartu: Eesti Keele Sihtasutus.

ISO 639 = *Code for the representation of the names of languages 2001* (http://www.oasis-open.org/cover/iso639a.html – August 10, 2004).

Jauhiainen, Marjatta 1998. *The type and motif index of Finnish belief legends and memorates: Revised and enlarged edition of Lauri Simonsuuri's Typen- und Motivverzeichnis der finnischen mythischen Sagen*. Folklore Fellows Communications 121 = 267. Helsinki: Suomalainen Tiedeakatemia.

Kippar, Pille 1986. *Estnische Tiermärchen: Typen- und Variantenverzeichnis*. Folklore Fellows Communications 102: 1 = 237. Helsinki: Suomalainen Tiedeakatemia.

Korb, Anu 2001. Suulise pärimuse kogumine oma rahvuskaaslaste juures [Collecting Oral Heritage among Kinsfolk]. *Paar sammukest: Eesti Kirjandusmuuseumi aastaraamat* 18. Tartu: Eesti Kirjandusmuuseum, pp. 164–193.

Krikmann, Arvo 1999. *Naljandite tüpoloogiast* [On the Typology of Folk Humour]. Tartu: EKM Folkloristika Osakond http://haldjas.folklore.ee/~kriku/HUUMOR/nljtypol.htm – August 10, 2004).

Kuperjanov, Andres. *Eesti kihelkonnad* [Estonian parisches] (http://www.folklore.ee/rl/folkte/kk/maps/we.htm – August 10, 2004).

Kuperjanov, Andres 2005. Does the Age of Multimedia change the World of Humanitarian Scholars? Ilmumas.

Kõiva, Mare 1990. *Estonskie zagovory: Struktura i zhanrovye osobennosti*. Doctoral paper. Manuskript. Tallinn: IEYaiL.

Kõivupuu, Marju 2001. Kogujalt kogujale [From Collector to Collector]. *Paar sammukest: Eesti Kirjandusmuuseumi aastaraamat* 18. Tartu: Eesti Kirjandusmuuseum, pp. 194–210.

Laugaste, Eduard & Normann, Erna 1999 [1959]. *Muistendid Kalevipojast* [Legends about Kalevipoeg. Estonian Legends: Giant and Hero Legends]. Eesti muistendid. Hiiu- ja vägilasmuistendid 1. Tallinn: Eesti Raamat. (http://www.folklore.ee/rl/folkte/myte/kalev/ – August 10, 2004).

LEPP = *Lõuna-Eesti pärimuse portaal* [South Estonian Folklore Portal] (http://www.haldjas.folklore.ee/lepp – August 10, 2004).

Loorits, Oskar 1999 [1939]. *Endis-Eesti elu-olu* I: *Lugemispalu kaluri ja meremehe* elust [Former Living in Estonia I: Readings About the Life of a Fisherman]. Tallinn: Kultuurkoondis & Tartu 1999 (http://www.folklore.ee/rl/pubte/ee/eluolu/elu1/saateks.html – August 10, 2004).

Pärdi, Heiki 2000. Mulgi müüt ja mulgi identiteet [The Myth and Identity of the Mulks – Inhabitants of the Viljandi County]. Kalle Gaston (toim). *Ettekandeid I mulgi konverentsilt: Tõrvas 23.10.99*. Mulgi Kultuuri Instituudi toimetised 1. Viljandi: Mulgi Kultuuri Instituut, pp. 15–31.

Remmel, Mariann 2002. Kohapärimuse digitaliseerimine Eesti Rahvaluule Arhiivis [Digitalisation of Local Lore in the Estonian Folklore Archives]. Hiiemäe, Mall (ed.). *Kogumisest uurimiseni: Artikleid Eesti Rahvaluule Arhiivi 75. aastapäevaks.* Eesti Rahvaluule Arhiivi toimetused = Commentationes Archivi Traditionum Popularium Estoniae 30. Tartu: Eesti Kirjandusmuuseum, pp. 115–133 (RFC 3066: Tags for the Identification of Languages, vt http://www.ietf.org/rfc/rfc3066.txt – August 10, 2004).

Saukas, Rein 2001. *Vanasõnade ja mõistatuste kogujad* [Collectors of Proverbs and Riddles]. Manuskript in Estonian Literature Museum.

Simonsuuri, Lauri l961. *Typen- und Motivverzeichnis der finnischen mythischen Sagen*. Folklore Fellows Communications 182. Helsinki: Suomen Tiedeakatemia..

Thompson, Stith 1961. *The types of the folktale: A classification and bibliography: Antti Aarne's Verzeichnis der Märchentypen*. Folklore Fellows Communications 75. Helsinki: Suoumen Tieteakademia.

Uther, Hans-Jörg 2004. *The Types of International Folktales: A Classification and Bibliography. Based on the System of Antti Aarne and Stith Thompson*. Folklore Fellows Communications 284–286. Helsinki: Suomen Tiedeakatemia.

Vabarna, Anne 2002a. Hagu, Paul (ed.). *Ale*. Tartu: EKM USN & Tartu Ülikool http://www.folklore.ee/rl/pubte/ee/setu/anne/ale/index.html – August 10, 2004).

Vabarna, Anne 2002b. Hagu, Paul (ed.). *Peko* II. Tartu: EKM USN & Tartu Ülikool (http://www.folklore.ee/rl/pubte/ee/setu/anne/peko2/index.html – August 10, 2004).

Voolaid, Piret 2002. *Traditsioonilisest ja improvisatsioonilisest eesti piltmõistatustes: Elektrooniline andmebaas* [On the Traditional and Improvisational in Estonian Droodles] (http://haldjas.folklore.ee/Reebus). Hiiemäe, Mall (comp.). *Kogumisest uurimiseni: artikleid Eesti Rahvaluule Arhiivi 75. aastapäevaks.* Eesti Rahvaluule Arhiivi toimetused = Commentationes Archivi Traditionum Popularium Estoniae 30. Tartu: Eesti Kirjandusmuuseum, pp. 141–156.

Voolaid, Piret 2003. Constructing Digital Databases of the Periphery of Estonian Riddles: Database Estonian Droodles. *Folklore: An Electronic Journal of Folklore* 25, pp. 87–92 (http://www.folklore.ee/folklore/vol25 – August 10, 2004).