EWA DAHLIG-TUREK
INSTITUTE OF ART, POLISH ACADEMY OF SCIENCES
ORCID 0000-0003-1895-8594

ANNA MARIA MATUSZEWSKA
INSTITUTE OF ART, POLISH ACADEMY OF SCIENCES
ORCID 0000-0002-3832-855X

CRAIG STUART SAPP
STANFORD UNIVERSITY
ORCID 0000-0001-8865-3177

THE ESAC COLLECTION: ORIGINS, NEW MUSICAL RESOURCES, AND TECHNICAL DEVELOPMENTS

ABSTRACT The EsAC (Essener Assoziativ Code), developed in the 1980s with tools for editing and analysis, is among the most important formats in digital musicology. Its core dataset, the *Essen Folksong Collection*, inspired many related projects worldwide. A key example is the Polish corpus of over 20,000 transcriptions from *Oskar Kolberg's Complete Works*, encoded in EsAC (2014–22) and later converted to Humdrum. The article outlines EsAC syntax, the Polish dataset, and new Verovio Humdrum Viewer features supporting both formats for in-depth analysis.

KEYWORDS computational musicology, EsAC, Humdrum, traditional music

ABSTRAKT *Kolekcja EsAC. Zarys historyczny genezy, nowe zasoby muzyczne i innowacyjne rozwiązania techniczne.* EsAC (Essener Assoziativ Code), opracowany w latach osiemdziesiątych wraz z narzędziami do edycji i analizy, należy do najważniejszych formatów w muzykologii cyfrowej. Jego główny zbiór, *Essen Folksong Collection*, zainspirował wiele podobnych projektów na całym świecie. Kluczowym przykładem jest polski korpus ponad 20 000 transkrypcji z *Dzieł wszystkich Oskara Kolberga*, zakodowany w EsAC (2014–22), a następnie przekonwertowany do formatu Humdrum. Artykuł omawia składnię EsAC, polski zbiór oraz nowe funkcje Verovio Humdrum Viewer, wspierające oba formaty w pogłębionej analizie.

SŁOWA KLUCZOWE muzykologia obliczeniowa, EsAC, Humdrum, muzyka tradycyjna

ISSN 0027-5344, e-ISSN 2720-7021

The use of computers for musicological research began in 1949 with Bertrand Bronson's project to analyse a folk song collection using IBM punch cards.[1] Developments in computational musicology were pioneered by digitisation projects such as those encoding the incipits of Italian Renaissance music,[2] sixteenth-century *frottola*,[3] almost all of Giovanni Pierluigi da Palestrina's works,[4] eighteenth-century symphonies[5] and compositions by Wolfgang Amadeus Mozart.[6] Although encoded incipits proved very effective in classifying works in musical collections and served as the basis for some analytical projects, it soon became clear that studies of compositional techniques or stylistic analysis could not be carried out solely on their basis. This led to the development of increasingly advanced languages for music coding, the functionality of which was also extended to analytical purposes (e.g. Digital Alternate Representation of Musical Scores[7] or Musical Information Retrieval[8]). And although many projects had to be abandoned owing to technical obstacles, numerous initiatives have survived to this day and continue to gain new iterations thanks to the rapid development of technology.

One such project is EsAC (Essener Assoziativ Code), developed in the 1980s by Helmut Schaffrath at the Universität-Hochschule Essen to encode monophonic music in a machine-readable format. EsAC was inspired by the Chinese notation JIANPU, which itself refers to the Galin-Paris-Chevé Method of the late nineteenth century, and also by the Krohn-Bartók system for folksong systematisation. Along with EsAC, a number of PC programs were developed to enable encoding control, melody display, and analysis. These programs still exist today; however, since they were designed for the DOS system, they are no longer convenient to use.

HISTORICAL BACKGROUND OF THE ESAC COLLECTION

In fields such as computational musicology, music information retrieval, and cognitive musicology, the collection popularly known as the *Essen Folksong Collection*,

1 Bertrand H. Bronson, 'Mechanical Help in the Study of Folk Song', *The Journal of American Folklore* 62 (1949) no. 244, pp. 81–86, doi.org/10.2307/536303.

2 Harry B. Lincoln, 'Some Criteria and Techniques for Developing Computerized Thematic Indices', in: *Elektronische Datenverarbeitung in der Musikwissenschaft*, ed. Hans Heckmann, Regensburg 1967, pp. 57–60.

3 Ibid., pp. 57–60.

4 Robert F. Erickson, 'Music and the Computer in the Sixties', in: *Proceedings of the May 5–7, 1970, Spring Joint Computer Conference*, Montvale 1970 (= AFIPS Conference Proceedings 36), pp. 281–285, doi.org/10.1145/1476936.1476983.

5 Jan Larue, Mary Rasmussen, 'Numerical Incipits for Thematic Catalogues', *Fontes Artis Musicae* 9 (1962) no. 2, pp. 72–75.

6 Hans-Rudolf Dürrenmatt, Murray Gould and Jan Larue, 'Die Notierung thematischer Incipits auf "Mark-Sense-Cards": Thematisch-bibliographische Katalogisierung im Zeitraum von 1600–1800 mit Hilfe elektronischer Datenverarbeitung', *Fontes Artis Musicae* 17 (1970) nos. 1–2, pp. 15–23.

7 R.F. Erickson, 'Music and the Computer in the Sixties', pp. 281–285.

8 Arthur Mendel, 'Some Preliminary Attempts at Computer-Assisted Style Analysis in Music', *Computers and the Humanities* 4 (1969) no. 1, pp. 41–52.

encoded in the EsAC format and later converted into Humdrum, holds remarkable significance. According to Google Scholar, citations referencing the Humdrum conversion reached 280 in 2025, reflecting its steady and ongoing growth in scholarly recognition.[9] However, due to the untimely death of its creator, Helmut Schaffrath, the precise history of its creation remains unclear. The original intent of the collection, the source data on which it is based, and other foundational details were not clearly documented, which has given rise to numerous myths. This text, however, focuses specifically on the early development of the EsAC format.

For many years, the EsAC collection was one of the most extensive digital repositories of monophonic music. The database used in musicological research conducted in 1991 by Barbara Jesser, a doctoral student of Professor Schaffrath, was initially named *Essener Liederdatenbanken* in German. This was later translated into English as the *Essen Folksong Collection* – a designation that has persisted as a commonly recognised name despite its relative inaccuracy, since neither the collection's original title nor the nature of the sources on which it was based, supports defining the entire repertoire as 'folk'.

The collection comprises over 6,250 monophonic tunes from various regions of Europe, with the largest subset consisting of approximately 5,300 tunes popular in the regions of Germany. In addition, the collection includes about 2,240 tunes from China.[10] Interestingly, most of the German tunes were not initially encoded in the EsAC format. Jesser's dissertation, which addresses the entire corpus of over 6,250 melodies, indicates that within the German subset 4,178 melodies were converted from the NUTOC format, created by Wolfram Steinbeck in 1982. Only 221 melodies, sourced directly from the Deutsches Volksliedarchiv in Freiburg, were encoded natively in EsAC.[11]

EsAC Software until 1990

Schaffrath and Jesser worked on a system of analytical procedures. As Jesser put it in her dissertation published in 1991, the aim was:

> to develop a system that makes it possible to store and analyse a large corpus of monophonic melodies and to keep it open for current research. […] The system is intended to provide assistance for classification, typologisation, style analysis and the search for variants in folk song research.[12]

9 Citation count retrieved from Google Scholar for: 'Helmut Schaffrath, "The Essen Folksong Collection," *Database* 6 (1995): 255', https://scholar.google.com, accessed 16 October 2025.

10 Helmut Schaffrath, *The Essen Folksong Collection in Kern Format*, ed. David Huron, Stanford 1995, https://kern.humdrum.org/cgi-bin/ksdata?file=README.txt&l=essen&format=kern, accessed 15 October 2025.

11 Barbara Jesser, *Interaktive Melodieanalyse: Methodik und Anwendung computergestützter Analyseverfahren in Musikethnologie und Volksliederforschung: Typologische Untersuchung der Balladensammlung des DVA*, Frankfurt am Main 1991, p. 111.

12 'Ziel dieser Arbeit ist, erstens ein System zu entwickeln, das es ermöglicht, einen großen Korpus einstimmiger Melodien zu speichern, zu analysieren und für aktuelle Recherchen offenzuhalten. […] Das System soll Hilfestellungen für Klassifikation, Typologisierung, Stilanalyse und Variantensuche in der Volksliedforschung geben', see B. Jesser, *Interaktive Melodieanalyse*, p. 2.

The methods developed in Jesser's doctoral thesis were designed for use with folk music from the Western cultural tradition.[13]

Beyond designing the EsAC encoding scheme, Schaffrath's aim extended to computer-based music analysis, which led to the development of several programs facilitating data manipulation and analysis. The software package MAPPET (Musik-Analyse, Play-in und Play-back for EsAC-Transkriptionen) was a set of five programs: PLAY, PIEPS, SYNTAX, ANA and TRANS.[14]

The first of these programs, PLAY, enabled the generation of input through a MIDI synthesiser. While melodies could alternatively be encoded as text files, Jesser stressed both the time-consuming nature of this process and its proneness to error. Melodies played on a computer-connected synthesiser were then converted into EsAC format. At the input stage, the user was required to specify the metre and the tonic (i.e., the reference pitch treated by the system as '1' with subsequent scale degrees defined in relation to it). The program interface featured a built-in metronome, and note durations were calculated indirectly from the pulses it generated. Melodies with a regular metre were divided into bars automatically, whereas in the case of irregular metres, bar lines, metadata, and phrase divisions had to be entered manually.

Error corrections could be made in the built-in editor, which displayed the EsAC code and allowed MIDI recordings to be played back for reference.[15] The program nevertheless had certain limitations: melodies could be encoded only within the range from C to b^3, and the smallest rhythmic value that could be entered was the demisemiquaver.[16]

The PIEPS program enabled users to listen to the encoded tunes through various sound generators, such as the PC sine-tone generator (PIEPST) or a synthesiser connected via a MIDI interface (PIEPSYN), in order to facilitate corrections. The latter provided an additional feature of accentuating strong beats in the metre, which made it easier to follow the progression of the tune. The synthesiser version proved particularly useful when working with large music collections, as it allowed for changes in register.[17]

The next program, SYNTAX, generated error reports, identifying issues such as incorrect encoding in the KEY field, improper use of symbols, faulty spacing, the presence of unidentified characters, inconsistencies in the total duration of the encoded bar, misdefined tonality, and other encoding errors.[18]

The ANA program was designed for analytical purposes, including: the creation of interval inventories (with indication of their direction), inventories of pitches and

13 Ibid.

14 Helmut Schaffrath to Eleanor Selfridge-Field, 'New Versions of ESTAFF and PAT', letter, Stanford: Center for Computer Assisted Research in the Humanities, 23 June 1991, https://esf.ccarh.org/ccarh-wiki/22-01_Mappet-Schaffrath.pdf, accessed 15 October 2025.

15 B. Jesser, *Interaktive Melodieanalyse*, pp. 132–135.

16 Ibid., p. 133.

17 Ibid., p. 136.

18 Ibid., p. 137.

durations, extraction of rhythmic patterns, analysis of scales and modes, calculation of ambitus and pitch contour, identification of phrase initiations and repetitions (determined by pitches or durations), and extraction of the melodic spine. Its output was saved in the AskSam database format (*.SAM).[19] By exporting the encoded files into a database, comparative analysis became possible even in the early stages of EsAC development. The procedure itself, however, differed considerably from how such analyses are carried out today.[20]

The last program, TRANS, was designed to facilitate the translation of data between two database formats: one employing language-dependent mnemonics (AskSam), and another based on a neutral line-numbering system (STAIRS). Conversion between these formats was possible in both directions, ensuring compatibility.[21]

EsAC software in 1990–1994

In 1990, Barbara Jesser withdrew from further cooperation, while Helmut Schaffrath concentrated on popularising the EsAC format, developing databases, and extending the analytical software. As part of this work, the EsAC software was expanded to include the five-module STRIP program, designed to reduce data unnecessary for subsequent analytical steps. From the information encoded in EsAC for analytical purposes, it was possible to: remove all but the data stored in the KEY and MEL fields (STRIP2), generate only the rhythmic information (STRIP3), extract melodic information without rhythmic values (STRIP4), or extract melodic information excluding repeated notes (STRIP5). The output data were saved in the AskSam format with a new extension, and the original version could be restored using the STRIP1 program.[22]

In the early 1990s, Schaffrath began collaborating with Ulrich Franzke from Bochum, who played a pivotal role in the final phase of the 'old' EsAC's development. Franzke created the PAT program, a tool designed to detect user-defined rhythmic–melodic patterns. The program could search both the melody database and the analytical results database, offering considerable flexibility: users could, for example, limit the search to the first or last line, the entire melody, or the results of the automatic analysis. Despite these innovations, the PAT program had notable limitations by today's standards. For instance, the database could process only up to 30 files, and users were restricted to a maximum of 30 search criteria.[23]

19 Damien Sagrillo, *Melodiegestalten im luxemburgischen Volkslied: Zur Anwendung computergestützter Verfahren bei der Klassifikation von Volksliedabschnitten*, Bonn 1999, vol. 1, p. 77.

20 For a very good illustration of the capabilities of the ANA program, see Helmut Schaffrath, 'The Essen Associative Code: A Code for Folksong Analysis', in: *Beyond MIDI: The Handbook of Musical Codes*, ed. Eleanor Selfridge-Field, Cambridge, MA 1997, pp. 343–361.

21 H. Schaffrath, letter to E. Selfridge-Field, 23 June 1991.

22 D. Sagrillo, *Melodiegestalten im luxemburgischen Volkslied*, pp. 82–83.

23 Ibid., pp. 83–90.

Among Franzke's other programs was Estaff, which translated EsAC syntax into traditional musical notation (*.PIC format). Its third version, ES3, could display both Western and Jianpu notation and allowed for the printing of encoded music. In addition, it was used for transposition and for displaying patterns generated through analytical procedures.[24]

During this period, several extensions were made to the original format, most notably the addition of a text field (TXT), in which lyrics were entered syllable by syllable, allowing songs to be displayed with text. Until Schaffrath's death in March 1994, numerous databases of varying size were created by individual students and scholars.[25] The total number of records encoded in the EsAC format during that period can be estimated at nearly 20,000, although no one possesses full knowledge of the actual figure, since these collections neither originated nor were compiled as a coherent resource. As they are linked only by the coding format, the terms 'EsAC repositories' or 'EsAC resources' seem the most appropriate.

After Helmut Schaffrath's death, Ewa Dahlig, his collaborator during the final period of his life, took charge of the project at the Institute of Art of the Polish Academy of Sciences (IA PAS) in Warsaw. Lacking both technical infrastructure and financial resources to engage IT specialists, the Polish team concentrated on developing databases of Polish traditional music. At the Institute, EsAC was used as an intermediate stage in the preparation of two volumes of the series *Polska Pieśń i Muzyka Ludowa* [Polish Folksong and Folk Music] edited by Ludwik Bielawski, dedicated to the traditional music of the various ethnomusicological regions of Poland.[26] The musical transcriptions for these volumes were first encoded in EsAC and only afterwards – through MIDI conversion – subjected to final editing in Finale.

The intended publishing objective was thus achieved, with the EsAC database of approximately 4,500 records as an additional benefit. Among other uses, it served as the basis for creating digital indexes of melodies: one in the form of a synoptic notation of all songs, reduced to a common abstract tonic and arranged in lexicographical order (by increasing pitch); and another presenting the material divided into systematic groups, clearly illustrating the affinity of songs within each category.

24 H. Schaffrath, 'The Essen Associative Code', p. 343–361.

25 For example: Damien Sagrillo – folksongs from Luxembourg (633) and Lorraine (500); Marco Piemontese – Italian folksongs (833); David Halperin – Hassidic tunes (112), Bach chorales (430); Ewa Dahlig – Polish folk tunes from Oskar Kolberg: *Collected Works*, vol. 1 (927).

26 *Kaszuby* [Kashubia], eds. Ludwik Bielawski, Aurelia Mioduchowska and Janina Szymańska, 3 vols., Warsaw 1997–98 (= Polska Pieśń i Muzyka Ludowa 2); *Warmia i Mazury* [Warmia and Masuria], eds. Barbara Krzyżaniak, Aleksander Pawlak and Monika Gruchmanowa, 5 vols., Warsaw 2002 (= Polska Pieśń i Muzyka Ludowa 3).

EsAC structure and syntax

The EsAC code was designed with the creation of musical databases in mind: its syntax is compact, and encoding symbols consist entirely of ASCII characters. An original EsAC record contains three obligatory fields:

(1) CUT [Title of the melody or text incipit],

(2) KEY [Line with encoding parameters],

(3) MEL [Encoded melodic notation].

Pitch, expressed by the numerals 1–7, is encoded relative to a reference tone represented by '1', which is always located in the middle octave – i.e., C1–B1 in the European pitch system and C4–B4 in the American system. In this way, EsAC captures the absolute pitch of the encoded sounds. Pitches above or below the principal octave are indicated by plus or minus signs preceding the numeral, with repeated signs marking additional octaves (e.g., '++1' for two octaves above the reference pitch). Numerals alone represent major or perfect intervals, which may be raised or lowered by appending (#) or (b) symbols (e.g., '3b' for a minor third above the reference pitch). Rests are denoted by the digit '0'. The base rhythmic value is expressed as a two-digit symbol, for example 08 for a quaver and 32 for a demisemiquaver.

The duration of tones without additional markings corresponds to the basic rhythmic value specified in the KEY field. An underscore symbol (_) following the pitch symbol doubles the rhythmic value, with each additional underscore mark having the same effect (e.g., two underscores quadruple the value, etc.). The dot (.) functions as an augmentation dot, as in standard music notation, increasing the duration by 50%. Triplets and other compound rhythms are encoded in parentheses.

Measures are separated by a blank space, and the end of the melody is marked with a double slash (//). Melodies are divided manually into phrases, with each phrase beginning on a separate line.

The interpretation of a tune's musical content begins with the KEY field. As shown in Fig. 1, the first parameter in the KEY field contains the database code for the tune; the second specifies the smallest rhythmic value (e.g., 16 for a semiquaver); the third indicates the tonic (F); and the fourth defines the metre (3/8).

There are also several optional fields in EsAC that store additional information when required, namely:

(1) TRD [Source specification, e.g. publication],

(2) REG [Region],

(3) FKT [Social function of the tune],

(4) BEM [Additional remarks],

with '[' and ']' marking the beginning and the end of a data field.

More recently, it has also become possible to generate EsAC data together with analytical features, which will be discussed in another section of this paper.

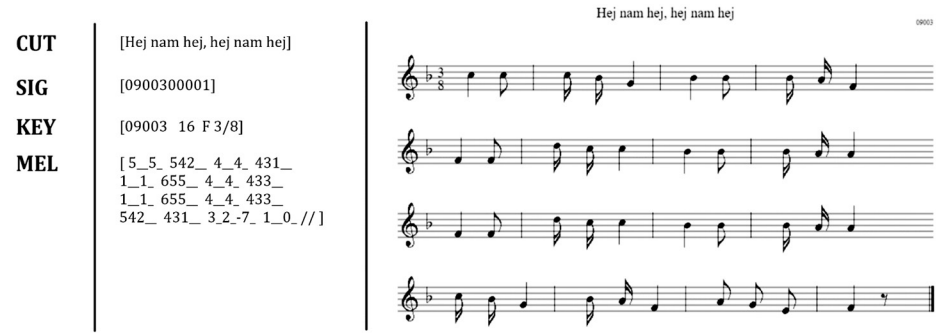


Fig. 1. 'Hej nam hej', a Polish traditional song encoded in EsAC and in standard music notation, Oskar Kolberg Collection, vol. 9, no. 003

Recent EsAC database and software development

In 2014, significant work was undertaken at IA PAS on the most important Polish source of traditional music: the monumental collection of musical records stemming from the field documentation of Oskar Kolberg (1814–90), the foremost collector of folklore in nineteenth-century Poland and one of the greatest in Europe. Kolberg, an educated musician and composer, spent fifty years collecting and successively publishing volumes devoted to all the major regions of the Polish lands. His oeuvre, embedded in a thorough and multifaceted description of folk culture, constitutes an inexhaustible treasury of knowledge on traditional music.

Kolberg aimed to document the entirety of the territory of the pre-partition (1772) Polish–Lithuanian Commonwealth. During his lifetime, he published thirty-three volumes, but he also left behind an extensive archive of materials that later served as the foundation for subsequent publications. The complete collection consists of eighty-six volumes[27] and encompasses twenty-five regions. Of these, sixty-six volumes contain musical material eligible for inclusion in the database.

As part of a project conducted between 2014 and 2022, more than 21,000 entries were encoded in the EsAC format on the basis of the printed editions. Plans are now underway to encode the volumes that were published during the course of the project.

The project involved a total of fifteen participants in the musicological (encoding) phase. The relatively high number of contributors resulted from the involvement of musicology students, for whom participation in the project – usually limited to encoding a relatively small number of records – also constituted their first experience

27 In total, there are 91 physical volumes, as vol. 59 comprises three parts, vol. 72 two parts, and vol. 73 three parts.

with digital musicology. In practice, a small core group of contributors played a crucial role in encoding Kolberg's materials. All musicology participants were trained in EsAC encoding by the project leader, Ewa Dahlig-Turek.

Each volume of Oskar Kolberg's *Complete Works* was encoded as a separate file containing records under a unified identifier, DWOK[+volume number] (e.g., DWOK09). In general, each melody was encoded as a single record. However, in cases where melodies contained segments with different key signatures or metres, or where instrumental and vocal parts were distinct, these sections were encoded as separate records. Similarly, when Kolberg documented variants of a melody's performance, each variant was encoded as an independent record. All records relating to the same melody are linked in the database through a shared component of their signature (SIG) field.

The material was encoded using plain text editors (*.txt files).[28] Each encoded entry underwent a second-person review to ensure quality. Owing to the varied levels of expertise among contributors, all material was pre-processed by Ewa Dahlig-Turek, who identified the tonic (reference pitch) and outlined the phrasing of each melody.

Within the framework of a project funded by the Polish Ministry of Science, it was possible not only to encode the Kolberg collection in EsAC,[29] but also to develop a new application for handling EsAC resources – webEsAC.[30]

This is a working application designed for interacting with EsAC resources. It contains a database of encoded volumes from Oskar Kolberg's *Complete Works* (DWOK)[31], but it also allows users to add their own datasets, over which they retain full editing rights.

The functionality of webEsAC includes:

– adding or removing record(s),

– editing a record,

– automatic re-encoding of a record (e.g., changing the tonic or the shortest rhythmic value),

– automatic addition of analytical information fields to each record,

– global 'find/replace' changes in any field,

– resource search (based on any record field),

28 This approach was initially necessary. Over time, as the webEsAC application evolved, it became possible to utilise its built-in editor and notation preview features.

29 'Tomy – DWOK' [volumes – DWOK], Oskar Kolberg, https://kolberg.ispan.pl/dwok/tomy/, accessed 20 September 2025.

30 The application was developed by a group of students from the Poznań University of Technology, working under the direction of Dr Ewa Łukasik: Joanna Sołomiewicz, Marcin Chmiel, Wojciech Błachowski and Jakub Kaczmarek, see Mariusz Chmiel et al., *Raport Techniczny* [Technical Report] RB-10/18: *WebEsAC – Internetowy system archiwizowania i wyszukiwania tradycyjnej muzyki polskiej w formacie EsAC* [*webEsAC* – An online system for archiving and retrieving traditional Polish music in the EsAC format], Poznań, University of Technology 2018.

31 In order to find it in the database, use the search path: Type → Name → DWOK.

– data filtering, with the option of combining multiple criteria,

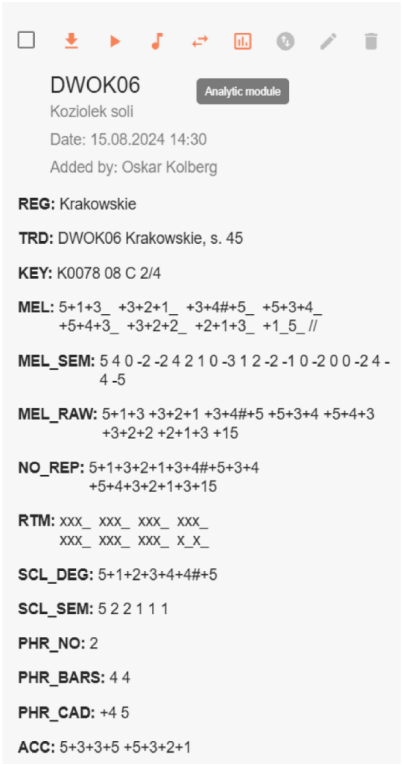
– audio playback of melodies (with adjustable tempo), courtesy of Ulrich Franzke,

– displaying musical notation,

– interactive dataset manipulation, including the selection of displayed information (i.e., fields),

– exporting automatically generated statistics on bar-level rhythmic structures and scales, for a selected range or the entire dataset, to facilitate further research with other software,

– exporting musical data in TXT or MIDI format for further research using other software.



MEL_SEM	melodic contour expressed as intervals (number of semitones)
MEL_RAW	melodic line without rhythmic values
NO_REP	melodic line without note repetitions
RTM	rhythmic representation
SCL_DEG	scale expressed as a sequence of successive degrees
SCL_SEM	scale expressed as a sequence of intervals (in semitones) between successive degrees
PHR_NO	number of phrases in the melody
PHR_BARS	number of bars in each phrase
PHR_CAD	cadential tones pf phrases
ACC	accentuated tones in the melody

Fig. 2. Overview of a single tune in webEsAC

Records are entered into the database either directly within the application or by importing text files prepared in external programmes, with one file corresponding to a single DWOK volume. All fields are fully searchable and may be used to filter the dataset by selected parameters, including searches for specific melodic or rhythmic patterns. Results, both original and filtered, can be exported for further analysis as required.

The structure of the record has been extended. Since the tune signature displayed in the KEY field contains an insufficient number of characters, an additional SIG field was introduced to allow for a more detailed definition of the record. In the past, analytical fields were generated by separate programmes of the MAPPET package. In webEsAC, they are automatically added to each record and display complete information, as shown in Fig. 2.

EXPANSION OF ESAC FUNCTIONALITY THROUGH CONVERSION TO HUMDRUM AND A NEW DATABASE

In 2024, Craig Stuart Sapp from Stanford University joined the project and introduced significant technical enhancements to the Verovio Humdrum Viewer (VHV)[32], enabling the following features:

– display and editing of files encoded in the EsAC format, including the loading of individual pieces or multiple melodies simultaneously,

– customisation of the filtering toolbar in VHV to process EsAC data. This is particularly useful for editing graphical layouts, such as:

– adjusting beams with the *autobeam* filter,

– displaying scale degrees below the musical notation (*deg*),

– visualising composite rhythms (*composite*).

Fig. 3 illustrates how the *autobeam*, *deg*, and *composite* filters work in the VHV editor for files encoded in the EsAC format.

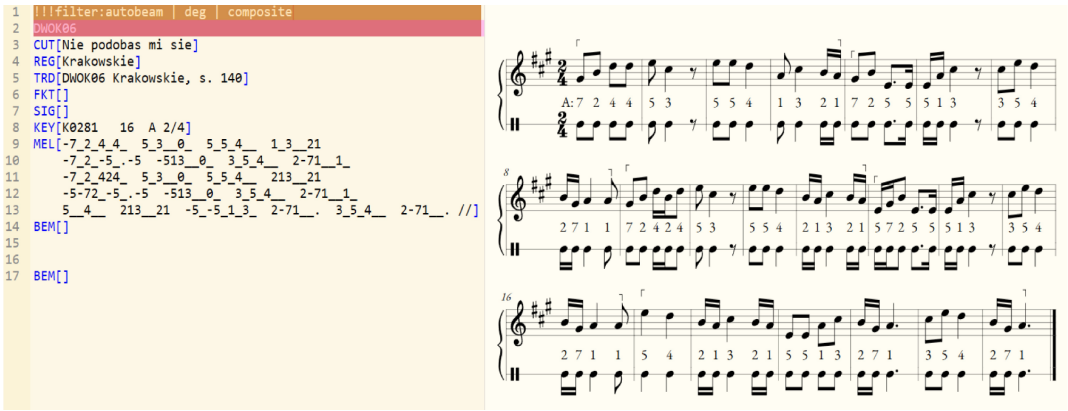


Fig. 3. 'Nie podoba mi się', a Polish folksong encoded in EsAC, Oskar Kolberg Collection, vol. 6, no. 281 with *autobeam*, *deg*, and *composite* filters

32 Verovio Humdrum Viewer, https://verovio.humdrum.org/, accessed 20 September 2025.

Sapp also introduced the *!!LO:LB:g=phrase* marker based on phrase tags in EsAC, which allows music notation in VHV to be organised by phrases. This feature enables phrases to be displayed on separate lines when the *Automatic line breaks* button is pressed in the navigation panel.

The most significant achievement, however, was the modernisation of the *esac2hum* converter,[33] which was essential for making 20,000 records from Oskar Kolberg's *Complete Works* accessible. Updates to the new converter included improved rendering of tuplets, such as triplets, enhanced handling of melodies with variable rhythm, ensuring proper division into measures.

The conversion of the corpus of Polish folk melodies is available in the ispan-esac Project on GitHub,[34] where two separate branches are provided: one containing the originally encoded EsAC files and the other their Humdrum conversions. All melodies are linked to their source materials when an EsAC file is converted to Humdrum. VHV users can then click on the PDF icon in the main toolbar to view the corresponding source material for a selected melody.

With the complexity of familiarising oneself with such an extensive musical database in mind, an interface for searching through Oskar Kolberg's collections was also designed,[35] incorporating elements of statistical musical analysis. The interface allows for the analysis of any three user-defined subcorpora as well as a detailed examination of individual melodies.

FUTURE PLANS OF ESAC DEVELOPMENT

The EsAC database currently contains approximately 40,000 tunes. However, despite the collection's vast and diverse nature, its research potential has not yet been fully realised. The future of the EsAC project seems to point in two main directions. On the one hand, given the simplicity of the code and the ease of use of webEsAC, which provides ready-made analytical results, it is worth continuing to develop databases in this format and to enhance the application itself by adding functionalities missing from the MAPPET package, such as melodic and rhythmic contour analysis and interval indices. MIDI-to-EsAC conversion would also be highly desirable. On the other hand, considering that the extensive EsAC resource is a valuable research material, it needs to be adapted to the extended capabilities offered by more ad-

33 Craig Stuart Sapp, *humlib: tool-esac2hum.cpp*, version 5746d3f, computer software, 2024, https://github.com/craigsapp/humlib/blob/master/src/tool-esac2hum.cpp, accessed 15 October 2025.

34 'Digital scores collected by Oskar Kolberg', GitHub repository, https://github.com/ispan-esac/kolberg, accessed 20 September 2025.

35 Anna Maria Matuszewska, 'Visual Presentation and Exploration of Musical Corpora: Case Study: Oskar Kolberg's Opera Omnia', in: *DLfM '23: Proceedings of the 10th International Conference on Digital Libraries for Musicology,* ed. Martha E. Thomae, New York 2023, pp. 82–85, doi.org/10.1145/3625135.3625148.

vanced software (e.g., Humdrum Toolkit, humlib, or HumdrumR). To achieve this, a labour-intensive addition of metadata will be required.

The collection has evolved not only with technological advances but, more importantly, in response to the changing needs of researchers. The main goal now is to standardise and modernise as many dispersed EsAC collections as possible by supplementing older data with missing information (e.g., ornaments, dynamic and agogic markings) and by adding song lyrics where applicable, which will greatly expand the scope of possible analyses. This entails going beyond the EsAC encoding format and preparing a more detailed edition in Humdrum, the collection's partner language.

Researchers will be able to move from basic rhythmic and melodic analysis to the examination of performance details, as well as to conducting text analyses and exploring text–music relationships in order to better understand the function of music in the context of folk culture.

The planned comprehensive analytical edition will undoubtedly be a highly valuable resource for educational purposes, such as learning to use digital music analysis tools. Above all, however, it will fulfill EsAC's original goal of making data available for music analysis, as envisioned by its creator, Helmut Schaffrath.

BIBLIOGRAPHY

Bronson, Bertrand H. 'Mechanical Help in the Study of Folk Song.' *The Journal of American Folklore* 62, no. 244 (1949): 81–86. doi.org/10.2307/536303.

Dürrenmatt, Hans-Rudolf, Murray Gould, and Jan Larue. 'Die Notierung thematischer Incipits auf "Mark-Sense-Cards": Thematisch-bibliographische Katalogisierung im Zeitraum von 1600–1800 mit Hilfe elektronischer Datenverarbeitung.' *Fontes Artis Musicae* 17, nos. 1–2 (1970): 15–23.

Erickson, Raymond F. 'Music and the Computer in the Sixties.' In: *Proceedings of the May 5–7, 1970, Spring Joint Computer Conference*, 281–285. Montvale: AFIPS Press, 1970 (= AFIPS Conference Proceedings 36). doi.org/10.1145/1476936.1476983.

Jesser, Barbara. *Interaktive Melodieanalyse: Methodik und Anwendung computergestützter Analyseverfahren in Musikethnologie und Volksliederforschung: Typologische Untersuchung der Balladensammlung des DVA.* Frankfurt am Main: Peter Lang, 1991.

Kaszuby, eds. Ludwik Bielawski, Aurelia Mioduchowska, and Janina Szymańska. 3 vols. Warsaw: Instytut Sztuki Polskiej Akademii Nauk, 1997–98 (= Polska Pieśń i Muzyka Ludowa 2).

Larue, Jan, and Mary Rasmussen. 'Numerical Incipits for Thematic Catalogues.' *Fontes Artis Musicae* 9, no. 2 (1962): 72–75.

Lincoln, Harry B. 'Some Criteria and Techniques for Developing Computerized Thematic Indices.' In: *Elektronische Datenverarbeitung in der Musikwissenschaft*, ed. Hans Heckmann, 57–60. Regensburg: Gustav Bosse Verlag, 1967.

Matuszewska, Anna Maria. 'Visual Presentation and Exploration of Musical Corpora: Case Study: Oskar Kolberg's Opera Omnia.' In: *DLfM '23: Proceedings of the 10th International Conference on Digital Libraries for Musicology*, ed. Martha E. Thomae, 82–85. New York: Association for Computing Machinery, 2023. doi.org/10.1145/3625135.3625148.

Mendel, Arthur. 'Some Preliminary Attempts at Computer-Assisted Style Analysis in Music.' *Computers and the Humanities* 4, no. 1 (1969): 41–52.

Sagrillo, Damien. *Melodiegestalten im luxemburgischen Volkslied: Zur Anwendung computergestützter Verfahren bei der Klassifikation von Volksliedabschnitten.* 2 vols. Bonn: Holos, 1999.

Schaffrath, Helmut. 'The Essen Associative Code: A Code for Folksong Analysis.' In: *Beyond MIDI: The Handbook of Musical Codes*, ed. Eleanor Selfridge-Field, 343–361. Cambridge, MA: MIT Press, 1997.

Schaffrath, Helmut. *The Essen Folksong Collection in Kern Format.* Data set, ed. David Huron. Stanford: Center for Computer Assisted Research in the Humanities, 1995. https://kern.humdrum.org/cgi-bin/ksdata?file=README.txt&l=essen&format=kern, accessed 15 October 2025.

Steinbeck, Wolfram. *Struktur und Ähnlichkeit: Methoden automatisierter Melodienanalyse.* Kassel: Bärenreiter, 1982.

Warmia i Mazury, eds. Barbara Krzyżaniak, Aleksander Pawlak, and Monika Gruchmanowa. 5 vols. Warsaw: Instytut Sztuki Polskiej Akademii Nauk, 2002 (= Polska Pieśń i Muzyka Ludowa 3).

KOLEKCJA ESAC. ZARYS HISTORYCZNY GENEZY, NOWE ZASOBY MUZYCZNE I INNOWACYJNE ROZWIĄZANIA TECHNICZNE

Muzyka tradycyjna należy do kluczowych obszarów eksperymentów muzykologii cyfrowej. Jednym z najważniejszych historycznie formatów kodowania zapisów muzycznych jest EsAC – Essener Assoziative Code, opracowany w latach osiemdziesiątych XX w. wraz z pakietem oprogramowania do edycji i analizy. Pierwszy stworzony w tym formacie zbiór, szeroko znany jako *Essen Folksong Collection* i obejmujący głównie melodie pieśni niemieckich, stał się impulsem do powstawania mniejszych kolekcji EsAC w wielu ośrodkach na świecie. Na tym tle szczególne znaczenie zyskuje polskie repozytorium ponad 20 000 zapisów muzycznych z *Dzieł Wszystkich Oskara Kolberga*, zakodowanych w l. 2014–22 w formacie EsAC, a następnie przekonwertowanych do formatu Humdrum. Artykuł omawia składnię, metody kodowania i narzędzia analityczne EsAC, przedstawia zwięzłą charakterystykę polskiego zasobu oraz najnowsze rozszerzenia funkcjonalności Verovio Humdrum Viewer, obsługujące oba formaty. Nadrzędnym celem tych działań jest udostępnienie cennego materiału wraz z towarzyszącymi narzędziami do pogłębionych, wieloaspektowych badań.

Ewa Dahlig-Turek, Anna Maria Matuszewska, Craig Stuart Sapp

Ewa Dahlig-Turek – ethnomusicologist, professor at the Institute of Art, Polish Academy of Sciences. Her research focuses on musical culture in Poland, particularly on the history of bowed chordophones and the morphology of 'Polish rhythms' in Polish and Scandinavian folk music. Following a research stay at Universität-Hochschule Essen (1992/93), she has been developing methods in the field of digital musicology. Since 1994, she has led the EsAC (Essener Assoziativ Code) project. In 2023, together with her research team, she completed the electronic edition of approximately 20,000 records from *The Complete Works of Oskar Kolberg* and co-developed the webEsAC application.
ewa.dahlig-turek@ispan.pl

Anna Maria Matuszewska – a researcher at the Institute of Art of the Polish Academy of Sciences, specialist in digital musicology. She obtained her doctorate as part of the Binationale Promotion programme implemented by the PAS Institute of Literary Research and the Hochschule für Musik Karlsruhe. As part of the MINIATURA grant, she studies user experience (UX) in the context of musicological tools, with a view to developing and testing a prototype of analytic interface supporting music researchers in their work.
anna.matuszewska@ispan.pl

Craig Stuart Sapp – a researcher at the Center for Computer Assisted Research in the Humanities (CCARH), operating under the auspices of Packard Humanities Institute, and a lecturer at Stanford University, where he earned a doctorate in computer music theory and acoustics. He collaborates with the Fryderyk Chopin Institute in Warsaw (NIFC) on digital projects, and is the technical director of several research initiatives, including the Josquin Research Project (josquin.stanford.edu), Tasso in Music Project (tassomusic.org), Polyrhythm Project (polyrhythm.humdrum.org), and The 1520s Project (1520s-project.org).
craig@ccrma.stanford.edu