Int J Digit Libr (2012) 12:105–119
DOI 10.1007/s00799-012-0086-z

The EthnoMuse digital library: conceptual representation and annotation of ethnomusicological materials

Gregor Strle · Matija Marolt

Published online: 22 March 2012

Abstract The paper presents two vital aspects of the EthnoMuse digital library. We first present the development of a flexible data model through FRBRoo and CIDOC CRM conceptualization of processes and relations in folk song and music realizations. The approach is novel in that it conceptualizes and integrates various folkloristic and ethnomusicological materials, and also standardizes the workflow of production and post-production processes related to recording and documenting of folk song and music. We also present how novel music information retrieval techniques were integrated into the library to provide support for annotation of its contents. Two case studies are presented: automatic segmentation and labeling of field recordings, and transcription of bell chiming recordings.

Keywords Digital libraries · Folk song and music · FRBRoo · CIDOC CRM · Music information retrieval

1 Introduction

The Institute of Ethnomusicology is the oldest institute of the Scientific Research Centre of the Slovene Academy of Sciences and Arts (SRC SASA); it was founded in 1934. The founding charter of the Institute includes among its basic functions the "compiling of as complete as possible a collection of Slovenian musical folklore". In line with basic folkloristic and ethnomusicological research, the main goals of the Institute are monitoring contemporary activities relating to folk music and dance traditions and professional intervention in those areas where folklorism draws on the heritage of folk song, music, and dance. Further, a vital part of research is focused on collection, preservation, research, and publication of gathered Slovenian folk song, music, and dance collections, and the development of the EthnoMuse digital library [1]. Since its founding, the Institute has built a large archive that includes content in various forms:

- folk song manuscripts with song transcripts (around 25,000 items dating from 1906 onwards).
- audio recordings of folk songs and melodies, with the oldest on wax cylinders from 1914 and around 30,000 field recordings on magnetic tape and digital media dating from 1955 onwards. The collections include detailed metadata of recording processes, transcriptions, places of origin, themes, first verses, beginnings of the melodies, meters, etc.
- visual documentation, including a photo archive and a collection of video documents.

The EthnoMuse digital library was developed to encompass the Institute's entire archive. It consists of a set of novel tools and techniques that assist with the digitization of the Institute's collections, support the ongoing production and post-production processes related to field recordings and documentation of Slovenian folk song, music, and dance, and enable manipulation and searching of the library's contents. In a wider context, an upshot of these activities is accessibility of parts of the EthnoMuse library to wider audiences and promotion of Slovenian cultural heritage.

Our primary motivation, related to the scope of this article is: (a) research and implementation of emerging technologies

G. Strle
Research Centre of the Slovenian Academy of the Sciences and Arts, Nov trg 2, 1000 Ljubljana, Slovenia
e-mail: gregor.strle@zrc-sazu.si

M. Marolt (✉)
Faculty of Computer and Information Science, University of Ljubljana, Trzaska 25, 1000 Ljubljana, Slovenia
e-mail: matija.marolt@fri.uni-lj.si

used in digital library environments—such as domain ontologies—for the purpose of integrating related cultural heritage collections and conceptual modeling of folkloristic and ethnomusicological domains; and (b) research and implementation of music information retrieval (MIR) methods and tools tailored to the specifics of folkloristic and ethnomusicological domains. In order to make the EthnoMuse valid as a basis and a tool for scientific research work at the Institute, specific concepts and methods that apply to research of folk song, music, and dance had to be implemented. This is due to the fact that scientists and researchers of the Institute are active contributors (creators of the archive) and primary users of the EthnoMuse library. On one hand, the nature of digitized content (folk song, music, and dance) demands implementation of specific folkloristic and ethnomusicological concepts and relationships into the underlying data model to enable its effective description and representation. On the other hand, these descriptions cannot be isolated from the production and post-production processes that individual items depend on; they cannot exist outside the events of their creation (collecting, recording, transcribing), documentation and archiving. They require a flexible data model enabling a formal structure for description of implicit and explicit concepts and relationships between concepts and events representing the production and post-production processes.

The paper is organized as follows. We start by an overview of international activities relevant to the presented research. Our motivation is to pinpoint a few examples and research activities related to folkloristic and ethnomusicological materials and frame our research in an international context. In Sect. 2, we propose a *flexible data model* for the EthnoMuse digital library that meets the previously presented requirements and show how the CIDOC Conceptual Reference Model[1] (CIDOC CRM) [2] and Functional requirements for bibliographic records object oriented (FRBRoo) definition [3] can be used to enhance the representational power of such a model. The fact that both ontologies serve as integration tools for information sources of cultural heritage domain, is an added value.

While the flexible data model provided a solution for structural and conceptual issues, the management of the EthnoMuse library also required the development of techniques for automatic manipulation and annotation of its content. The third part of the paper presents some applications for access, manipulation, and management of musical contents in the EthnoMuse. These applications are built on top of *flexible data model* and employ novel MIR techniques to assist the data entry and annotation process. With audio recordings being the largest and most important unstructured documents in the EthnoMuse, we focused our work on developing techniques for automatic extraction of semantic descriptors from audio recordings and present examples of two such applications in Sect. 4.

1.1 Context

When assessing the functionality of a digital library, we first need to investigate the needs and the demands of its end-users. The key assumption is that the problem domain and people involved must be understood before tools and methods can be developed and later assessed [4], especially in an environment where the library represents the primary source for scientific research. Thus, to be effective, methods and tools presented throughout this article are tailored to the needs and the demands of our primary users—scientists and researchers of the SRC SASA. The EthnoMuse digital library reflects the environment specific to the folkloristic and ethnomusicological studies of Slovenian cultural heritage, i.e., research methods, tools, and conceptual organization used in explorations of folk song and music and related phenomena are fully incorporated into the EthnoMuse library system. As such, EthnoMuse is a small, but very specialized digital library with specific tools and methods for annotation, manipulation, and organization of content, and access limited to scientists and researchers of respective fields. Thus, finding comparable projects that focus on research and application of novel information retrieval techniques to folk song and music research is a difficult task, for many reasons: (a) digital library projects are usually parts of larger initiatives and oriented towards covering wide end-user populations and large quantities of different types of content (a good example with strong focus on music materials is *Variations 3* [5]); (b) as a consequence, the information retrieval tools used by such initiatives are too general to be relevant for the specifics of folkloristics and ethnomusicological research; (c) there are numerous smaller projects that focus on research and application of MIR techniques, with most of the research geared towards application of MIR methods to contemporary and classical music [6,7]. Though frequently present as test collections, folk song and music have rarely been part of serious introspection by the MIR community. As [8] observe: "... impediments for fruitful collaboration are the unfamiliarity of researchers in both fields with each other's methods and traditions, and the non-formalised nature of many [*folk song and music*] *concepts and theories*". The latter presents a serious problem both for conceptual modeling of folk song and music and for successful application of MIR methods. One of the reasons lies in the oral tradition of folk song and music itself—with its ever-changing, evolving and dissolving variants and variant families, and strong intertextuality, these aspects of folk song

[1] CIDOC CRM model was being developed from 1996 under the auspices of the ICOM–CIDOC (International Council for Museums–International Committee on Documentation) Documentation Standards Working Group. The definition of the CIDOC CRM model has now become ISO standard 21127 [2]

and music are hard to grasp. Thus, it might be necessary to develop MIR algorithms that represent music at a higher, more conceptual abstraction level than the level of notes [7]. A further problem with folk song and music is that it does not always correspond to the western concepts that underlie the currently available content-based methods [9]. While paramount to folk song and music research, the discussion and application of these concepts in the MIR community has generally been ignored (for discussion see [7–12]).

On the other hand, from the perspective of digital library research, there is an obvious lack of applications of formal conceptual models or domain ontologies to the folkloristic and ethnomusicological domain [13]. Domain ontologies (such as CIDOC CRM and FRBRoo) are necessary for integration of information sources across various collections of cultural heritage and an important step towards formal conceptual representation of concepts, relationships, and contextual information specific to the domain. As such, they represent formal semantics for cultural heritage information and promise to overcome the interoperability issues on lower (metadata) level [14–16]. In the paper, we argue for the necessity of such ontologies in folkloristic and ethnomusicological domains, and show how a formal conceptual representation of processes related to production and post-production of folk song and music has been applied to the EthnoMuse digital library. With this in mind, we present an overview of four digital library initiatives, related to our work.

The WITCHCRAFT project [17] sets as its objective to develop a fully functional content-based retrieval system for folk song melodies stored as audio and notation, building on the best practices of MIR research. More precisely, the project's aim is the design, implementation, and evaluation of a melody search engine that is capable of handling large amounts of audio and notation data and retrieves ranked lists of melodies based on similarity measure(s) that reflect the musical characteristics of the repertoire. The system's potential is demonstrated by integration into the The Dutch Song Database [18] (Nederlandse Liederenbank) of the Meertens Instituut, which contains more than 125,000 songs in the Dutch and Flemish language, from the Middle Ages through the twentieth century. A follow-up project entitled Dutch songs on Line [19] plans to expand into a comprehensive new database by adding a new, large data set with full transcriptions of lyrics, links to scans of the sources and contextual information. WITCHCRAFT is comparable to the presented work in the way it uses MIR techniques specifically developed for the domain of folk song and music; we are also testing its very promising melody search system based on similarity of folk song variants [10] for integration into the EthnoMuse library.

Danish Folklore Archive's material is based on over 200 fieldtrips made by Evald Tang Kristensen from 1867 to 1910, during which he collected ca. 250,000 stories, songs, games, rhymes, cures, observations of daily life, etc, from around 6,500 named informants [20]. The search engine offers different methods, e.g., text searching through various parameters of music books and/or melodies, and a separate interface for melodic search based on code types (incipit or accented), 8-digit melody code, number of matches between input code and desired variants, and time signature. Both incipit and accented code approaches are very effective for searching through folk song and melody variants. On the other hand, there is no support for MIR analysis of audio content, as the archive's tools are limited to content-based searching through symbolically represented melodies. Also, there is no information on the use of domain ontologies.

Digital archive of Finnish Folk Tunes [21] holds around 9,000 tunes and their relevant details (notation, key, meter, place of collection, lyrics, collector, etc.). The activities in the digital archive of Finnish Folk Tunes cover the development of applications, such as various musical search and query methods (e.g., query-by-humming or finding melodies using symbolic information), comparative research on musical features and lyrics, development of tools for computational analysis and availability of the archive for research purposes. Access to collections is available through a search engine that enables tunes to be searched by keyword, collection, musical type, geographical location, and musical content, while visualization of results includes maps describing geographical distributions of the songs and other musical details [22]. The archive is freely available for research purposes and offers tools for basic research on the musical structure, classification, and other such large-scale analytic and empirical techniques [23] for symbolically represented melodies. There is no information on the use of domain ontologies.

The Indiana University Digital Library Program is extending the *Variations* digital music library system [5] to incorporate the FRBR conceptual model. Variations is a large digital music library software system that provides online access to streaming audio and scanned score images with a flexible access control framework to ensure respect for intellectual property. In addition to access tools, variations also includes analysis and annotation tools useful in music teaching and learning. These tools are being developed to cover a wide variety of musical styles and demands of various user populations and are not specifically targeted to folk music. Relevant to this study is a new, ongoing project entitled *Variations as a Testbed for the FRBR Conceptual Model* [24]. The project is focused on testing the FRBR conceptual model in a real-world environment, and on providing data, code, and system design specifications that can be reused by others interested in FRBR [25–27]. The "real-world environment" here means primarily MARC environments and transformation/mapping of MARC records onto FRBR structure.

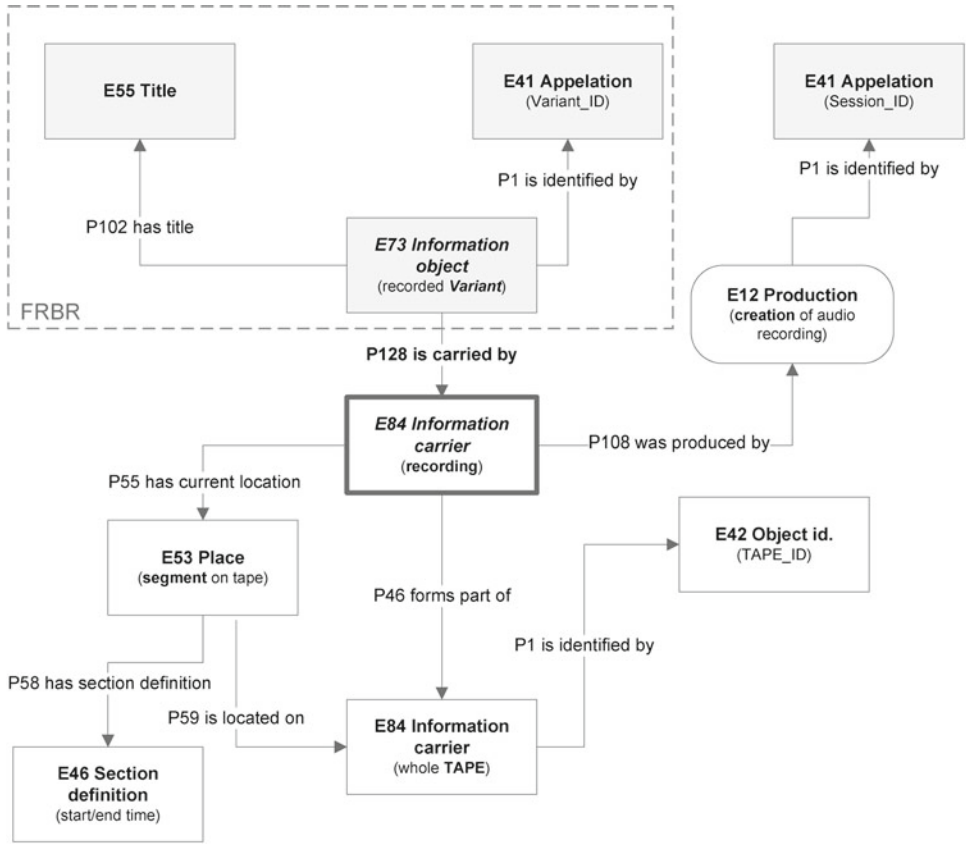

Fig. 1 Mapping a recording session in CIDOC CRM

2 The flexible data model[2]

It is in the nature of a folk song to be performed and mediated verbally. According to Le Boeuf [28], this has some major consequences:

> Non-performing arts products are physical objects that carry conceptual objects; such physical objects are the basic "documents" that are preserved in a collection and described in a catalogue; through such documents, the conceptual objects they carry are preserved and described as well; it is also possible to gather "documentation" about them, their creation, and the way they were perceived.
>
> Performing arts products are events that convey conceptual objects; there is therefore no basic "document", but only some "documentation", which, consequently, becomes primordial, as the conceptual object conveyed by a show can be somewhat preserved and described only by preserving and describing that documentation.

Audio recordings remain the basic scientific source for folkloristic research [29] and a core feature of folk song representation. Since there is no document, the representation of a particular folk song in a recorded variant is only the representation of its physical characteristics—audio documentation of the event of singing, not (of) the event/performance itself. Because of this strong interdependence, the concept of *event* represents a core feature in conceptual design of the proposed *flexible data model*. It serves as an anchor between a recorded variant, a person/actor involved (performer, recording team, etc.), and place and time of the recording session. Based on the CIDOC CRM domain ontology for cultural heritage information, the concept of event enables the representation and integration of various production (i.e., field recording session) and post-production processes through time. Figure 1 shows the mapping of the recording session (creation of an audio recording) in CIDOC CRM [30].

While the conceptual structure of CIDOC CRM proved to be efficient for modeling events, such as production and

[2] This section presents the creation of the *flexible data model* for conceptual representation of production and post-production processes within the EthnoMuse environment. It is important to note that, while the final version of the *flexible data model* is FRBRoo-ized and thus formal, the process of creating the data model has been progressive: the project ran parallel with the development of the FRBRoo ontology by the International Working Group on FRBR/CIDOC CRM Harmonization, spanning from the publication of the FRBRoo draft in 2006 to the latest revision of FRBRoo v1.0.1 in January, 2010. Since the first stable version of FRBRoo has not been available until 2008, we have in-between evaluated some other options for conceptual modeling, such as the use of the formal CIDOC CRM and later some kind of non-formal combination of CIDOC CRM and FRBR. Shortcomings of each of these solutions are presented in the following text.

post-production processes, it does not support the representation of complex folkloristic and ethnomusicological descriptions of folk song and music *variants* and their material realizations. In the *flexible data model*, entities and relationships that form between the performed folk song and its recorded variant (as the embodiment of the folk song's immaterial nature in the event of recording) are far too complex to be represented only by CIDOC CRM *E73 Information object* and *E84 Information carrier* (Fig. 1). While CIDOC CRM proved as insufficient for representing complex bibliographic relationships of folk song and music, their numerous variants and carriers, FRBR does just that.

2.1 FRBR and FRBRoo

FRBRoo [3] has its roots in FRBR model [31]. The FRBR model was originally designed as an entity-relationship model by a study group appointed by the International Federation of Library Associations and Institutions (IFLA) during the period 1991–1997, and was published in 1998. The original entity-relationship definition of FRBR is referred to hereafter as FRBRer.

While names FRBRer and FRBRoo suggest difference of method from entity-relational to object oriented modeling, both conceptual and structural changes become evident in the attempt of their application. Here, we shortly discuss major differences between both models.

In the FRBRer model, every new performance creates a new variant that belongs to a particular folk song variant type—*Work* [31]. Following the FRBRer structure, this *Work* is represented by *Expression*, embodied in *Manifestation* (recorded variant) and carried by an *Item* (audio carrier). Such a framework represents the hierarchical structure and description of content on various levels.

The representation of the data model shown in Fig. 2 is a non-formal combination of two conceptual models: FRBRer and CIDOC CRM, with the aim to enable:

1. representation of entities and relationships describing folk song and music, based upon FRBRer, especially 1st group entities [31]: *Work*, *Expression*, *Manifestation*, and *Item*; and
2. representation of production and post-production processes based on the CIDOC CRM concept of *Event*.

As Fig. 2 shows, the production of an individual *audio recording* and description of the recorded *variant* are linked with the overall *documentation* (field record) of a *Production Event*.

This way, not only is each individual recording or post-production session associated with other recording sessions through Production Event but also FRBRer entities describing recorded variants are associated. The four building blocks of the FRBRer model represent different levels of abstraction and complexity of bibliographic information: (a) *Work* is the product of an intellectual process, the sum of concepts which "appear in the course of the coherent evolution of an original idea into one or more expressions that are dominated by the original idea" [3]. *Expressions*, on the other hand are "the intellectual or artistic realisations of works in the form of identifiable immaterial objects, such as texts, poems, jokes, musical or choreographic notations, movement pattern, sound pattern, images, multimedia objects, or any combination of such forms that have objectively recognisable structures. The substance of Expression is signs. Expressions cannot exist without a physical carrier, but do not depend on a specific physical carrier and can exist on one or more carriers simultaneously. Carriers may include human memory." [3]. *Manifestation* is the physical embodiment of an *Expression* of a *Work*—it "represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form", while *Item* is a single exemplar of a Manifestation [31]. Each Manifestation is indirectly linked with the Production Event (individual recording session) through the Item (audio recording)—the information carrier of the recorded variant produced in a particular production event. Consequently, performers (singers, musicians, etc.) belong to the Manifestation level (of recorded variant), while recording team, transcribers, collectors belong to the Production Event level (particular recording session).

While offering a more detailed (bibliographic) representation as compared to the CIDOC CRM representation from Fig. 1, the combination of FRBRer and CIDOC CRM represented in Fig. 2 does not express any formal definition of any of the two models. The transition from Fig. 1 to 2 depicts the lack of descriptive power of CIDOC CRM representation in Fig. 1, where the model obviously lacks abstraction and complexity of bibliographic information. On the other hand, the combination of both models in Fig. 2 is non-formal. The problem is two-fold: (1) such combination (Fig. 2) is inconsistent with conceptual definitions and logical structure of classes and properties (and relationships between them) of individual models (FRBRer or CIDOC CRM), and consequently (2) the integrative and mediative power of each individual model is thus lost. In a complex digital library system with various kinds of collections and materials (folk song and music, dance, customs, etc.), the need for formal integration of both models is vital.

The initiative of formal integration of both models is represented in the current version of FRBRoo v.1.0.1 [3], prepared by the International Working Group on FRBR/CIDOC CRM Harmonization. "The harmonization between the two models is also an opportunity to extend the scope of the CIDOC CRM to bibliographic information, which paves the way for extensions to other domains and formats, such as EAD, TEI, MPEG7, just to name a few. Consequently, it also extends

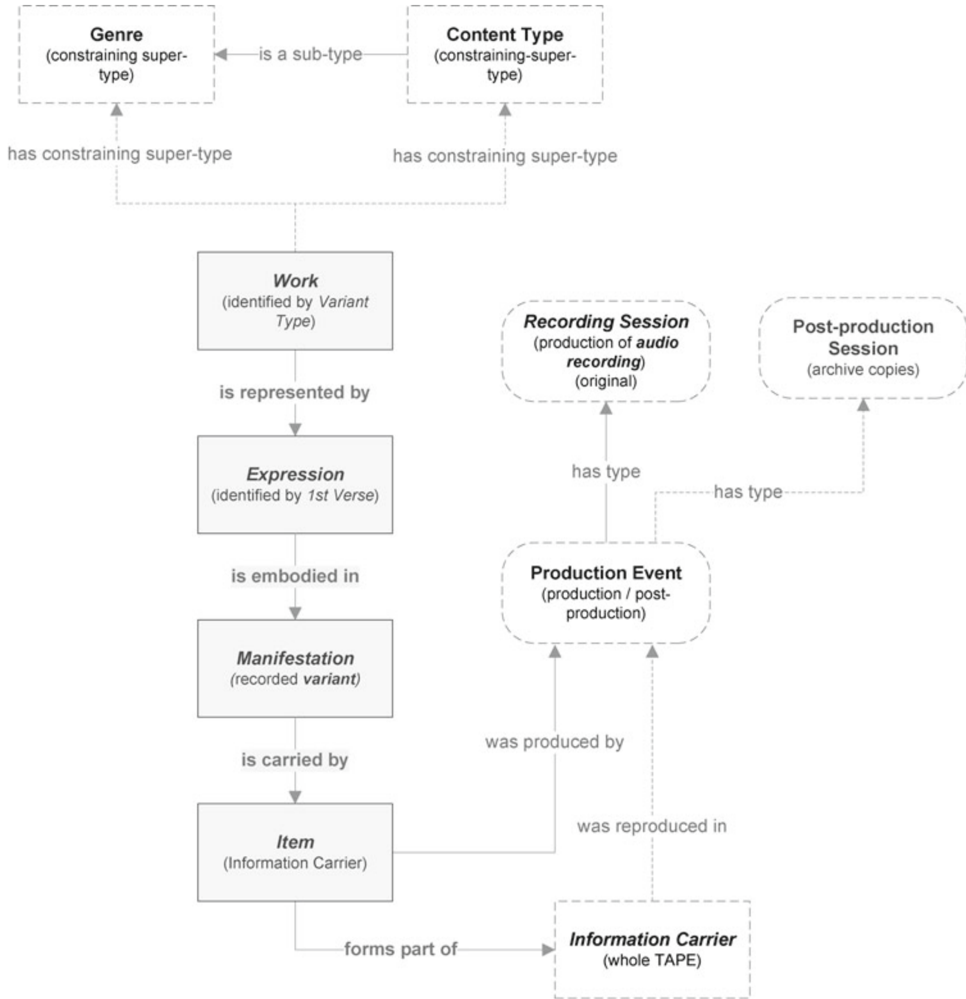

Fig. 2 Applying basic FRBRer entities (*continuous line*) to the CIDOC CRM structure

the scope of FRBR to cultural materials, since FRBR inherits all concepts of the CIDOC CRM, and opens the way for FRBR to benefit from further extensions of the scope of CIDOC CRM, such as the scientific heritage of observations and experiments. . . . Besides, CIDOC CRM is explicitly compatible in formalism with the World Wide Web Consortium's Resource Description Framework (RDF), which can only be beneficial for FRBR" [3].

2.2 Mapping to FRBRoo: changes

In the attempt to map the existing data model onto FRBRoo, some conceptual and structural changes had to be made, especially due to the way FRBRoo interprets *Manifestation*. While in FRBRer *Manifestation* can be interpreted as a representation of either material or immaterial (an *Item* or *Information Object*), in FRBRoo *Manifestation* is split: ". . . into two distinct classes, corresponding to the two possible ways of interpreting the ambiguous definition provided for Manifestation in FRBRer, namely F3 Manifestation Product Type and F4 Manifestation Singleton. Whereas F3 Manifestation Product Type is declared as a subclass of the CIDOC CRM class E55 Type, and therefore as a subclass, too, of the CIDOC CRM class E28 Conceptual Object (a merely abstract notion), F4 Manifestation Singleton is declared as a subclass of the CIDOC CRM class E24 Physical Man-Made Thing, and therefore as a subclass, too, of the CIDOC CRM class E18 Physical Thing." [31]

Following above, the *F3 Manifestation Product Type* represents the immaterial level (a CIDOC CRM subclass of *E28 Conceptual Object*), while *F4 Manifestation Singleton* represents the material level (a CIDOC CRM subclass of *E18 Physical Thing*).

In the old data model (Fig. 2) the distinction between an audio recording (*Manifestation*) and concrete information carrier (*Item*) is kept. This follows the initial FRBRer interpretation of *Manifestation* and *Expression* as ". . . the set of concepts expressed in one particular set of signs, independently of the materialisation of that set of signs. . ." [3]. Consequently, the *Expression* level conveys different realizations or "communicative" forms of a particular folk song variant—sung, instrumental, or spoken, whereas their materialization is modeled on the *Manifestation* level (in form of an audio recording, melody and/or text transcription etc.). Individual information carriers with their unique IDs are represented on *Item* level. In FRBRoo the original audio recording is defined as unique (as a "manuscript"), thus *F4 Manifestation Singleton* overlaps with an *Item*.

However, mapping of the data model onto FRBRoo solved some inconsistencies of the FRBRer model (from Fig. 2). First, the creation of a folk song and music variant is now modeled simultaneously with its materialization—the production of the original audio recording. As already mentioned, an audio recording represents the most authentic documentation of a particular variant, while other forms of variant representations such as text and melody transcriptions are considered as variant's partial realizations. Further, the production of an audio recording and the production of a transcription represent two different events, the latter usually based on the analysis of the former. The new flexible data model (based on FRBRoo) enabled these conceptual distinctions as shown in Fig. 3. Such a combination arose mainly due to the existing audio-archiving practices, where documentation and preservation of audio recordings on various carriers, as well as techniques and procedures used in production and post-production events, is of greatest importance. Figure 3 shows the FRBRoo structure of the flexible data model[3] based on the representation of folk song variant and its material realizations: production of *audio recording* and individual *transcriptions* of *melody* and *text*. Individual *Variant Type* is defined as *F15 Complex Work*, a *conceptual whole connecting* all folk song variants based on 1st Verse; it is further constrained by *Genre* and *Content Type* (both represented by CIDOC CRM *E55 Type*). Each individual variant is represented as *F15 Complex Work* itself—through individual (*F14 Individual Work*) and complex works it connects different forms of its unique realizations (as shown in Fig. 3).

The relationships between the *variant*, *audio recording*, and *transcription* of melody and text are made explicit: the *variant* is represented by *audio recording* (*F26 Recording*) while transcriptions of melody and text are both instances of *F22 Self-contained Expression*. The latter are only realizable through transcription of *audio recording* and are therefore in the existing model represented as derivations. This relationship is defined by *F15 Complex Work* (*melody/text*)—*R2 is derivative of*—*F15 Complex Work* (the *variant*), while *F15 Complex Work* (*melody/text*)—*R2.1 has type*—*E55 Type* defines the type of *transcription*.

Audio recording is the primary and most representative expression of a *variant*, and separated from the variant's other (particular) realizations. The relationship between the variant (*F15 Complex Work*) and its audio recording (*F26 Recording*) realized in the recording session (*F29 Recording Event*) is additionally strengthened by *R40 has representative expression*, defining audio recording as "... the most characteristic expression of the instance of F15 Complex Work of which it is an expression" (FRBRoo, p. 135). And further, *F26 Recording*—*R42B has representative manifestation singleton*—*F4 Manifestation Singleton* (original).

Another characteristic of FRBRoo are events. For example, the production event *F29 Recording Event* is simultaneously the event of expression creation (F26 Recording) and the event of production of the *F4 Manifestation Singleton*. "Although F2 Expression is an abstract entity, a conceptual object, the creation of an expression inevitably also affects the physical world ...The spatio-temporal circumstances under which the expression is created are necessarily the same spatio-temporal circumstances under which the first F4 Manifestation Singleton is produced" [3].

Post-production events (for example the making of archival copies) is represented by the *F33 Reproduction Event*, making the relationship between original recording and a copy explicit by *R31B has reproduction*. Figure 3 further shows relationships between FRBRoo and CIDOC CRM ontologies: *F4 Manifestation Singleton* is an instance of CIDOC CRM *E84 Information carrier*, while all events (for example *F29 Recording Event* or *F33 Reproduction Event*) are subcategories of CIDOC CRM *E12 Production*. Mapping of FRBRoo events onto the CIDOC CRM conceptual structure is vital in EthnoMuse: it enables integration of individual elements of folk song and music into context, both in relation to individual events (time and place), actors (performer, recording team, transcriber, etc.), and environment (social and cultural characteristics of event), which are essential for folkloristic and ethnomusicological research. Further, the representation of context on different (immaterial and material) levels of variant's realizations presents a first step towards semantic representation of content. From the implementation point of view, the FRBRoo model inherits from CIDOC CRM[4]: "it is an ontology in the sense used in computer science. It has been expressed as an object-oriented semantic model, in the hope that this formulation will be comprehensible to both documentation experts and information scientists alike, while at the same time being readily converted to machine-readable formats such as RDF Schema, KIF, DAML+OIL, OWL, STEP, etc. It can be implemented in any Relational or object-oriented schema. CRM instances can also be encoded in RDF, XML, DAML+OIL, OWL and

[3] Mappings to FRBRoo and CIDOC CRM are represented by rectangular boxes for *classes* and oval boxes for *events*; CIDOC CRM *classes*, *events* and *relationships* are represented by dashed borders and lines and lighter color area, while basic building blocks of FRBRoo stucture are bordered in bold.

[4] Detailed instruction and tools for mapping can be found on http://www.cidoc-crm.org/crm_mappings.html

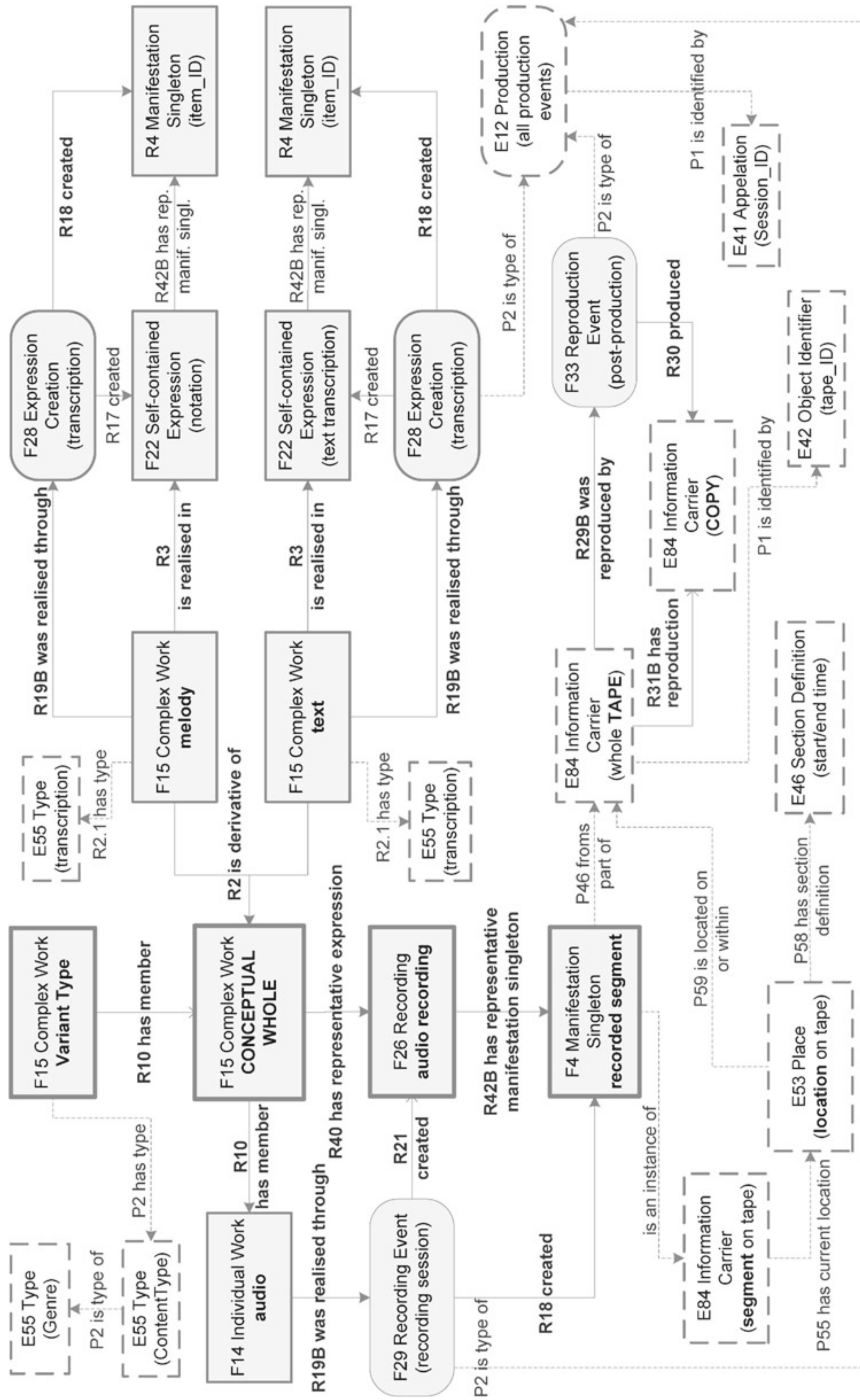

Fig. 3 The flexible data model—FRBRoo (image is rotated for clarity)

others" [2]. In general, the FRBRoo model promises great opportunities for creating retrieval systems that better support user information seeking and is being widely embraced by library communities, prompting expectations that future cataloging standards and, consequently, library practice and system development will undergo major changes [14,25,32–35]. Thus, apart from ensuring strong data modeling foundation and a prospect of information integration, FRBRoo plays an important role in user information seeking by explicit representation of different levels of abstraction and complexity of bibliographic information [3,31,36].

3 EthnoMuse for the end-user

The data model presented in the previous section represents the core upon which a set of tools for users of the EthnoMuse digital library was built. We first developed the *EthnoMuse Librarian* as a desktop application that integrates a set of tools for storage, annotation, and manipulation of digitized contents and its metadata and is primarily intended for use by researchers at the Institute. To enable flexible browsing and searching of the EthnoMuse collection we developed the *EthnoMuse Explorer* as a web application that can be used by a variety of users over the internet. Applications were built with Microsoft.NET and SQL Server development tools, which represent a modern development platform that is well proven in the software industry. This gave us a set of good development tools, as well as a very flexible platform for future enhancements.

3.1 The EthnoMuse Librarian

The Librarian is primarily intended to be used as an administrative tool for maintaining, adding, and annotating contents in the EthnoMuse library. A view of its user interface is shown in Fig. 4. It provides a platform that supports easy manipulation of all content types within the library: manuscripts, music scores, audio recordings, images, and videos.

Manuscripts are the oldest items in the library. These are handwritten documents of a song's lyrics and musical notation with accompanying metadata (performers, dates, etc.). The handwriting used is far too difficult for current optical music recognition and optical character recognition algorithms, so manuscripts were first scanned and then manually transcribed. Sibelius [37] was used to input music notation, because its easy to use features were well appreciated by researchers at the Institute. Scores are stored in two formats: Sibelius' native format enables easy score visualization and playback in desktop and web applications (with Sibelius' Scorch plugin), while MusicXML [38] provides a platform independent score representation and is used for further parsing and score processing by the EthnoMuse Explorer and

Fig. 4 Screenshot of the EthnoMuse Librarian: a scanned manuscript, its transcription as well as some of the metadata describing the item are shown

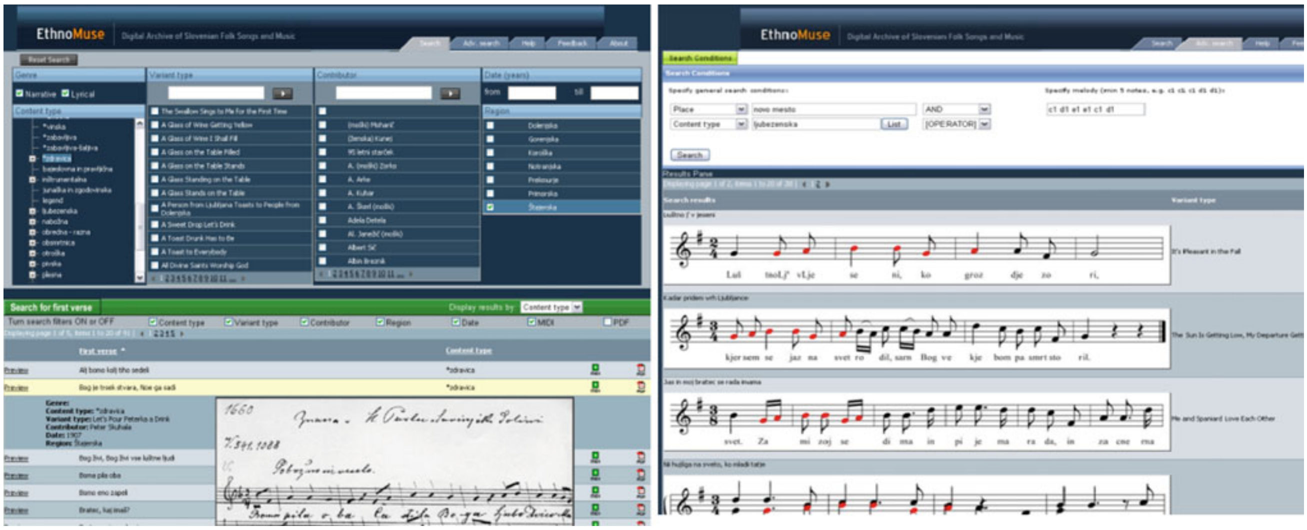

Fig. 5 The EthnoMuse Explorer—the browser interface is shown on the *left*, the search interface on the *right*

other applications. Next to scores and scans of original manuscripts, metadata ranging from musical properties such as meter, tempo, and melodic structure to genre, variant and content types, geographical characteristics and performers are also stored in the library.

Field recordings represent the largest part of the archive's contents. Researchers are still recording new materials, so support was added for archival as well as collection and annotation of new recorded materials. As recordings are also being published on CDs and used in various other projects, support for post-production processes is available, which enables researchers to trace a recording from its initial capture over all post-production steps (copying, editing, mastering . . .) to its final issued form. In addition to audio, field recordings are accompanied by a technical report that contains recording details, such as equipment and microphone placement, and is also stored in the library. The digitized audio of a recording session resides on external storage and is linked to the database record of a recording session by a URI. Field recordings available for listening and annotation are converted from their original digitized form to 128 kbps mp3 streams for faster access over the network. To segment a recording into individual units (e.g., speech, song, etc.), we developed a special annotation tool that includes an automatic segmentation and labeling algorithm (see Sect. 4.1) and is used to assist researchers with the annotation process.

The Librarian application also supports storage and manipulation of visual documents, such as images, videos of folk dances and Labanotation scores.

3.2 The EthnoMuse Explorer

While the Librarian is mostly intended to be used by researchers at the Institute, browsing and searching (parts of) the EthnoMuse library is made available to a larger community. For this purpose, we developed a web interface to the library, called the EthnoMuse Explorer, which allows for browsing and searching of the library's contents. The browser enables users to browse the collection by selecting a combination of genres, content types, variant types, performers, dates and regions and displaying songs belonging to the chosen criteria (Fig. 5, left). The search interface (Fig. 5, right) supports Boolean queries that can combine full-text searches over textual attributes with melody-based searches. The latter are implemented with standard n-gram search techniques [39], allowing for approximate matching, as well as with similarity measures based on similarity of folk song variants [10,40]. Visualization of the found melodies is provided with the on-the-fly lilypond [41] score rendering that also highlights the searched melody within each found score.

4 MIR and the EthnoMuse library

To enable easier access to the large number of audio recordings in the EthnoMuse library and to alleviate some of the tedious work needed to manually annotate these recordings, as well as perform meaningful searches within audio data, we are continuously developing new algorithms and tools that apply the latest findings of MIR researches onto the folk music domain. We focus our work on algorithms for automatic extraction of semantic descriptors from audio—a very time-consuming and difficult task to perform manually. In this section, we present two applications that use such algorithms and that are already integrated into the EthnoMuse library.

4.1 Automatic segmentation and labeling of field recordings

Ethnomusicological field recordings are recordings made “in the field,” capturing music in its natural habitat. Starting in the early twentieth century and continuing to the present day, ethnomusicologists and folklorists have travelled and made recordings in various parts of the world primarily to not only preserve folk music but also make it available for further researches, such as studies of acculturation and change in music, comparative studies of music cultures and studies of the music making process and its effect through performance. Segmentation of field recordings into meaningful units, such as speech, sung, or instrumental parts, is one of the first tasks researchers face when a recording is first being studied. It is also a prerequisite for further automatic processing, such as extraction of keywords, melodies, and other semantic descriptors.

Segmentation of audio recordings has been extensively explored for applications such as speech recognition (removal of non-speech parts, speaker change detection), segmentation in broadcast news or broadcast monitoring. Approaches to segmentation include either first classifying short periods of the signal into desired classes using some set of features and then making the segmentation or first finding change points in features and forming segments and later classifying the segments [42–44]. Authors use a variety of features, classifiers, and distances depending on the nature of signals to be segmented.

The algorithm incorporated into the EthnoMuse digital library was designed to robustly label and segment ethnomusicological field recordings into consistent units, such as speech, sung, and instrumental parts [45]. Resulting segmentations are comparable to manual segmentations researchers make when studying recordings. Field recordings are documents of entire recording sessions and typically contain interviews with performers intertwined with actual performances. The latter include singing (solo or group), reciting, instrumental pieces, as well as bell chiming, which is a Slovenian folk tradition of playing rhythmic patterns on church bells. The quality of recordings varies a lot and depends on their age, equipment used, location (inside, outside), and type of event (arranged recording session or recording of a public event). As these are live recordings of amateur folk musicians, they usually contain lots of “noise” and interruptions, such as silence when performers momentarily forget parts of songs, false starts and restarts, laughter, dancing noises, interruptions by other persons, dogs barking or cars driving by. Performances may also change character; singing may become reciting, a second voice may join or drop out of a performance, etc.

The described nature of field recordings calls for a robust segmentation algorithm that does not over-segment a recording at each interruption—for example, we are not interested in each boundary separating speech and sung parts, as only some of them are actual segment boundaries. We also need to distinguish between several different classes of signals and last, we are not interested in millisecond-exact segment boundaries or exact labeling of each small recording fragment; sometimes placing a boundary between two performances is a very soft decision and accuracy of a few seconds is good enough. Taking these points into account, we have developed a three-step probabilistic algorithm for segmentation.

First, a standard classification algorithm (logistic regression) is used to classify short audio segments into a set of predefined classes. Features extracted from these segments (including means and variances of MFCCs, spectral entropy, and signal energy) were classified into five classes: speech, solo singing, choir singing (any performance with two or more voices belongs to this class), instrumental (including instrumental with singing), and bell chiming.

After classification, a set of candidate segment boundaries is obtained by observing how the energy and class distribution change, and finally the recording is segmented with a probabilistic model that maximizes the posterior probability of segments given a set of candidate segment boundaries with their probabilities and prior knowledge of lengths of segments belonging to different classes. For more details of the algorithm, see [45].

We evaluated the algorithm on a set of field recordings from the library and compared them with manual segmentations made by researchers at the Institute. Overall, on our test set of 30 field recordings containing 840 segments, the algorithm reaches an accuracy of 80 % precision and recall on the task of finding correct segment boundaries, and 86 % precision on classifying the segments into one of the five classes. The classification confusion matrix is given in Table 1.

Most of the errors made by the classification algorithm are easy to explain. The confusion of speech and solo singing segments is understandable, if we take into account that singers are not professional musicians, they are often old persons and their singing close to reciting or very monotonous. Confusion between solo and choir singing occurs in choir segments sung in unison, as well as duet singing, while instrumental and bell chiming segments are correctly classified in most cases with confusion mostly arising between the two classes. Overall, segmentation accuracy is good enough to provide valuable information to the annotator.

4.1.1 Integration with the EthnoMuse

We integrated the segmentation algorithm into the EthnoMuse Librarian by developing an audio annotator that visually displays segments and their labels. We color-coded the five labels (speech—dark blue, solo singing—light blue, choir singing—green, instrumental—orange, bell

Table 1 Confusion matrix of the classification algorithm

	classified as				
	Speech (%)	Solo singing (%)	Choir singing (%)	Instrumental (%)	Bell chiming (%)
Speech	79	14	4	3	0
Solo singing	13	61	24	1	1
Choir singing	2	10	82	3	3
Instrumental	1	3	3	82	11
Bell chiming	0	0	2	7	91

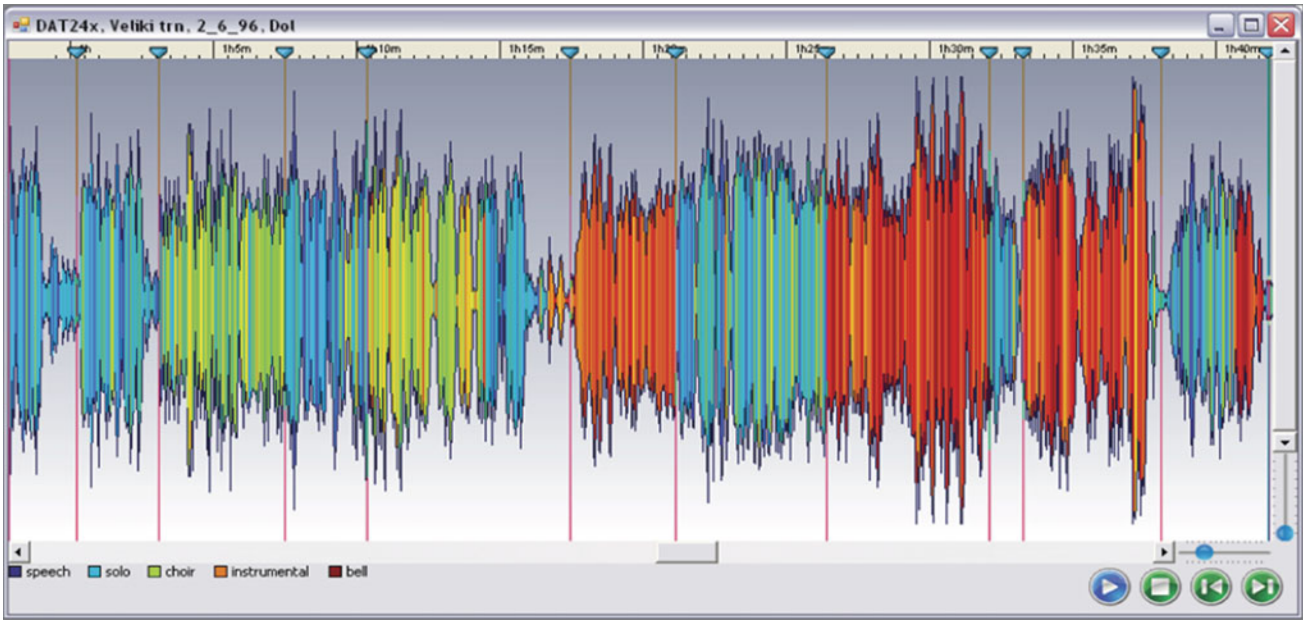

Fig. 6 Visualization of automatic labeling and segmentation of field recordings

chiming—dark red) and calculated the color of each classified short audio segment by interpolating between these colors while taking the segment's classification probabilities over the five classes into account. The colors were chosen so that more similar classes (e.g., speech and solo) have more similar colors. Figure 6 displays visualization of a waveform of a 45-min long field recording. We can immediately discern the seven solo singing segments (light blue), as well as two choir segments (green–yellow) and four instrumental segments (orange–red). Vertical lines indicate segment boundaries found by the probabilistic segmentation algorithms. The audio annotator tool enables users to listen and edit the automatically induced annotations and input further metadata for each segment.

To enable easier preview of contents of entire field recordings, we also generate a thumbnail preview of each recording that summarizes its contents. The preview is generated by first clustering the classification probabilities of individual short segments into five classes using k-means clustering and then generating the thumbnail by calculating the color of each cluster center and taking the cardinality of each cluster into account to obtain the correct cluster proportions. An example is given in Fig. 7, which shows thumbnails of ten field recordings. From these thumbnails, we can immediately discern the contents of each recording; for example, one can see that T525-0301 mostly contains solo singing, while T522-0101 mostly contains instrumental and choir music, with bits of solo and speech.

4.2 Enabling access to bell chiming recordings with transcription

Bell chiming is a Slovenian folk music tradition that still exists in its original context today. It takes place in the church tower and its original role is strongly connected to Christian religious contexts. Bell chiming combines the signaling, ritual, and musical functions, because it is most often used to call the faithful to mass in a musical way, and at the same time to mark important church holidays [46].

Slovenian-style bell chiming is performed by the musicians holding the clapper and striking the rim of the stationary bell at regular intervals. The sound is thus not produced

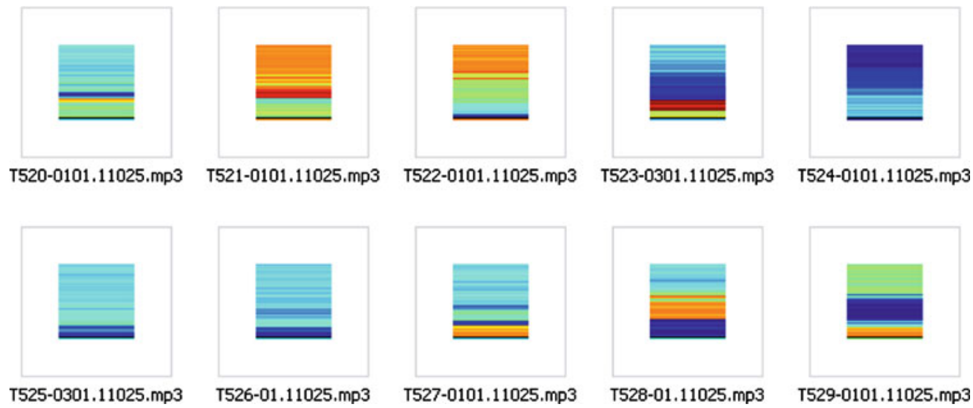

Fig. 7 Visualization of field recording contents. *Colors* indicate the proportions of different types of music in each recording. (Color figure online)

by a swinging bell hitting the clapper, but by the clapper, typically held close to the rim, hitting the bell's rim. This gives musicians more control in altering the rhythm, speed, dynamics, and accents of individual strikes, as well as leaving out strikes if desired. As a rule, each musician is responsible for playing one bell, and should strike the bell only with its clapper (touching the bell's rim with hands or other tools is not allowed). Another important rule in bell chiming is that two tones can never be played at the same time, but exceptions do occur.

Bell chiming tunes contrast one another in the method of playing, the number of bells used, and their rhythmic and metric structure. Tunes themselves consist of repeated rhythmic patterns into which various changes, typically dynamic and agogic are included to enliven the performance [46].

The EthnoMuse digital library holds a large collection of bell chiming recordings, collected from the 1950s onwards and only small parts of the archive have been manually transcribed and annotated. To alleviate the need for time-consuming manual transcription of these recordings, we developed an algorithm for automatic transcription of such recordings. Transcription processes an audio file and outputs a symbolic representation which can be used when searching the collection for a specific bell chiming pattern, thus enabling easier access to these recordings. Automatic music transcription is a difficult problem to solve, although methods are improving constantly; Klapuri and Davy [47] provide an extensive overview of the current state of the art. For transcription of bell chiming recordings, we developed a two step algorithm [48]. The algorithm first analyzes a bell chiming recording and estimates the number of church bells and their approximate spectra using prior knowledge of church bell acoustics and bell chiming performance rules. It then uses non-negative matrix factorization with selective sparsity constraints and a complex-domain onset detection algorithm to perform the transcription.

To test the algorithm, we manually transcribed a set of bell chiming recordings containing three to five bells and evaluated the number of correctly transcribed bells. The average precision and recall are at around 78 % on a test collection containing 905 notes. The result is good if we consider that we are transcribing real field recordings, which are affected by factors such as poor microphone placement, weather conditions, bell tower acoustics, etc. Even though onset detection itself works very well, most of the errors are made due to indistinctive bell onsets, which may occur because of the aforementioned factors, long bell decay times or change of dynamics by performers. We designed a tool to assist researchers with transcription by displaying the calculated transcription and spectra of the found bells, as well as giving researchers the possibility to correct transcription errors and play back parts of the signal and/or transcription. Each transcribed recording can be saved to a MIDI file, which is then used to index the obtained transcriptions and make them available within the EthnoMuse Explorer (Fig. 8).

5 Conclusion

The flexible data model is at the core of the EthnoMuse digital library. It represents a formal (FRBRoo and CIDOC CRM) conceptualization of ethnomusicological materials in the library, making implicit relationships between a folk song variant and its numerous realizations in production and post-production processes explicit, thus assisting folkloristic and ethnomusicological research. With its contextual representation of content it is an important aid in the development of search and retrieval tools and methods. Further, it is a basis for existing and future developments of applications and tools for management and manipulation of materials in the library. It was proven to be a consistent model for describing existing and integrating new multimedia materials, as well as offering means for possible integration and mediation of similar collections from other libraries.

The presented EthnoMuse tools integrate novel MIR algorithms to facilitate the work of researchers using and maintaining the library, as well as providing other users with extended browsing and searching capabilities. Our future

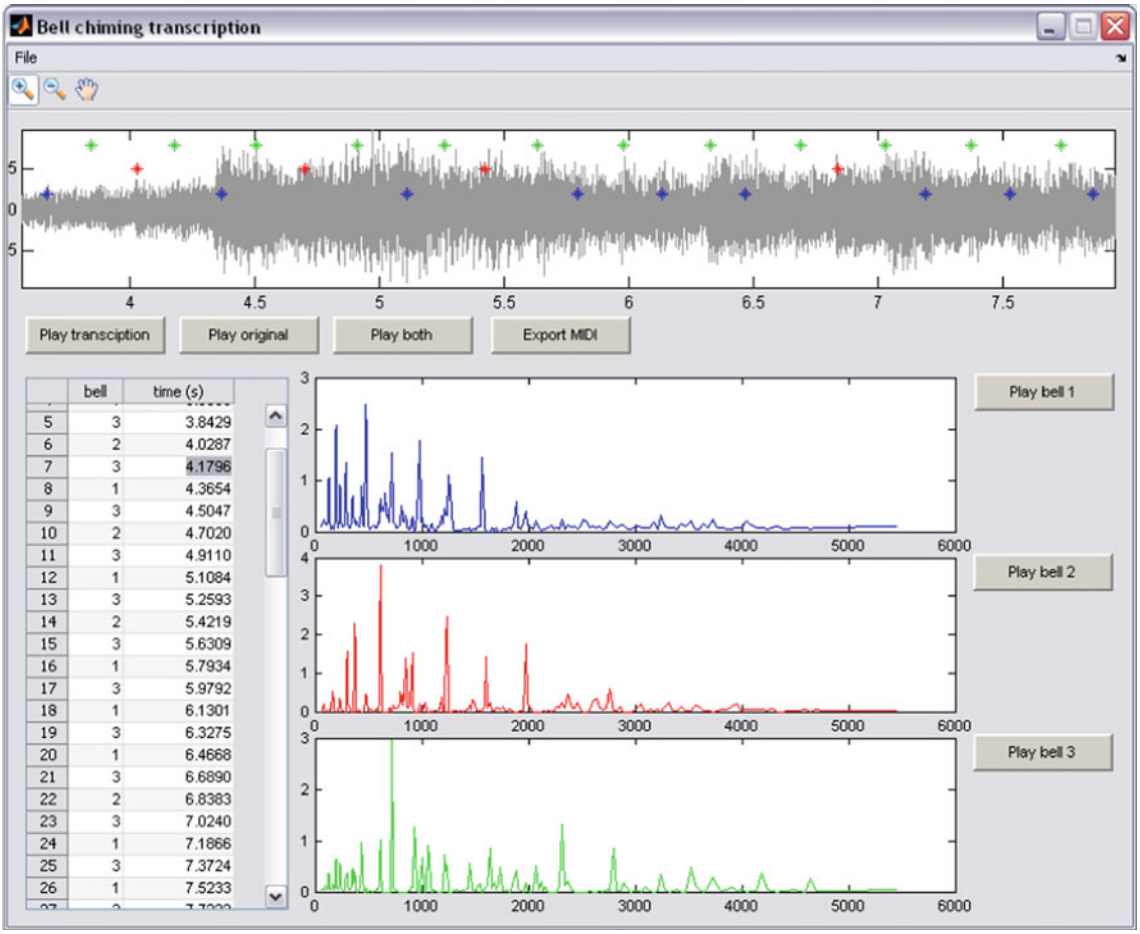

Fig. 8 Bell chiming transcription tool

plans include further development of new MIR algorithms and tools to assist researchers with time-consuming manual tasks, such as estimation of tempo, meter, and instruments played, as well as to study similarities and relations between folk song variants and provide visualizations of the library's contents.

Acknowledgments This work was supported in part by the Slovenian Government-Founded R&D project EthnoCatalogue: creating semantic descriptions of Slovene folk song and music.

References

1. EthnoMuse: multimedia digital archive of Slovenian folk, music and folk dance culture. Slovenian Government Founded R&D Project (2006–2008)
2. ICOM/CIDOC CRM Special Interest Group: Definition of the CIDOC conceptual reference model: version 5.0.1. ICS-FORTH, Heraklion, Greece (2009)
3. International Working Group on FRBR and CIDOC CRM Harmonisation: FRBR object-oriented definition (draft version 1.0.1) (2010)
4. Seadle, M.: Project ethnography: an anthropological approach to assessing digital library services. Library Trends **49**, 370–385 (2000)
5. Variations3: The Indiana University Digital Music Library. Retrieved from http://www.dlib.indiana.edu/projects/variations3/
6. Casey, M.A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., Slaney, M.: Content-based music information retrieval: current directions and future challenges. Proc. IEEE **96**(4), 668–696 (2008)
7. Typke, R., Wiering, F., Veltkamp, R.C.: A survey of music information retrieval systems. In: Proceedings of the International Conference on Music Information Retrieval, London, UK, pp. 153–160 (2005)
8. van Kranenburg, P., Garbers, J., Volk, A., Wiering, F., Grijp, L., Veltkamp R.C.: Towards integration of MIR and folk song research. In: Proceedings of ISMIR, Vienna, Austria, pp. 505–508 (2007)
9. Cornelis, O., Lesaffre, M., Moelants, D., Leman, M.: Access to ethnic music: advances and perspectives in content-based music information retrieval. Signal Process. **90**, 1008–1031 (2010)
10. van Kranenburg, P. Garbers, J., Volk, A., Wiering F., Veltkamp, R.C.: Musical models for folk-song melody alignment. In: Proceedings of ISMIR, Kobe, Japan, pp. 507–512 (2009)
11. Moelants, D., Cornelis, O., Leman, M., Gansemans, J., De Caluwe, R., De Tré, G., Matthé, T., Hallez, A.: The problems and opportunities of content-based analysis and description of ethnic music. Int. J. Intang. Herit. **2**, 57–68 (2007)
12. Van Kranenburg, P., Garbers, J., Volk, A., Wiering, F., Grijp, L.P., Veltkamp, R.C.: Collaboration perspectives for folk song research

and music information retrieval: the indispensable role of computational musicology. J. Interdiscip. Music Stud. **4**, 17–43 (2010)
13. Nicolas, Y.: Folklore requirements for bibliographic records: Oral traditions and FRBR. Cat. Class. Q. **39**, 179–195 (2005)
14. Doerr, M., Hunter, J., Lagoze, C.: Towards a core ontology for information integration. J. Digit. Inf. **4**, 169 (2003)
15. Doerr, M.: The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI Magazine **24** (2003)
16. Staab, S., Studer, R. (eds.): Handbook on Ontologies. Springer, Berlin (2009)
17. WITCHCRAFT - What Is Topical in Cultural Heritage: Content-based Retrieval Among Folksong Tunes (2010). Retrieved from http://www.cs.uu.nl/research/projects/witchcraft/
18. The Dutch Song Database—Nederlandse Liederenbank (2010) Retrieved from http://www.liederenbank.nl/index.php?wc=true
19. Dutch Songs on Line (2010) Retrieved from http://meertens.musipedia.org
20. The Danish Folklore Archives—Dansk Folkemindesamling (2010) Retrieved from http://www.dafos.dk/
21. Eerola, T., Toiviainen, P.: Digital archive of Finnish folk tunes. Computer Database, University of Jyvaskyla (2004)
22. Toiviainen, P., Eerola, T.: Visualization in comparative music research. In: COMPSTAT 2006, Proceedings in Computational Statistics, pp. 209–219 (2006)
23. Eerola, T., Toiviainen, P.: Mir in Matlab: The midi toolbox. In: Proceedings of the International Conference on Music Information Retrieval, Barcelona, Spain, pp. 22–27 (2004)
24. Variations as a Testbed for the FRBR Conceptual Model (2010) Retrieved from http://www.dlib.indiana.edu/projects/vfrbr/
25. Byrd, J., Charbonneau, G., Charbonneau, M., Courtney, A., Johnson, E., Leonard, K., Morrison, A., Mudge, S., O'Bryan, A., Opasik, S. et al.: A white paper on the future of cataloguing at Indiana University (2006). Retrieved from http://www.iub.edu/libtserv/pub/Future_of_Cataloging_White_Paper.pdf
26. Riley, J., Mullin, C., Colvard, C., Berry, A.: Definition of a FRBR-Based Metadata Model for the Indiana University Variations3 Project. Phase 2: FRBR Group 2&3 Entities and FRAD (2008)
27. Riley, J., Mullin, C., Hunter, C.: Automatically batch loading metadata from MARC into a work-based metadata model for music. Cat. Class. Q. **47**, 519–543 (2009)
28. Le Boeuf, P.: "... That struts and frets his hour upon the stage and then is heard no more.": the elements that should be accounted for in a conceptual model for performing arts and the information relating to their archives. In: Workshop at the Centre de documentation de la musique contemporaine, Paris, France (2006)
29. Schüller, D.: Methodik und technik der phonographischen Feldforschung. In: Deutsch, W., Walcher (eds.) Sommerakademie Volks-kultur, Volksliedwerk, Wien (1994)
30. Stead, S., Strle, G.: CIDOC CRM and Folksong. In: Clark, J., Hagemeister, E. (eds.) Computer Applications and Quantitative Methods in Archaeology Proceedings of the 34th Conference, pp. 445–449, Fargo, United States (2006)
31. IFLA Study Group on the functional requirements for bibliographic records: functional requirements for bibliographic records: final report. K.G. Saur, Munich, Germany (1998)
32. Antelman, K., Lynema, E., Pace, A.K.: Toward a 21st century library catalog. Inf. Technol. Libr. **25**, 128–139 (2006)
33. Feng, L., Jeusfeld, M.A., Hoppenbrouwers, J.: Beyond information searching and browsing: acquiring knowledge from digital libraries. Inf. Proces. Manag. **41**, 97–120 (2005)
34. Mann, T.: The changing nature of the catalog and its integration with other discovery tools. A Critical Review. J. Libr. Metadata **8**, 169–197 (2008)
35. Matheson, S., Davidson, S.: The evolution of providing access to information: is the online catalog nearing extinction?. Legal Ref. Serv. Q. **26**, 57–89 (2007)
36. Raimond, Y., Sandler, M.: Using the semantic web for enhanced audio experiences. In: Audio Engineering Society Convention, New York, NY, USA (2007)
37. Sibelius score annotation software (2010). Retrieved from http://www.sibelius.com
38. Good, M., et al.: MusicXML: an internet-friendly format for sheet music. In: XML Conference Proceedings, Orlando, Florida (2001)
39. Downie, J.S.: Evaluating a simple approach to music information retrieval: conceiving melodic n-grams as text, Ph.D. Thesis, University of Western Ontario (1999)
40. Bohak, C., Marolt, M.: Calculating similarity of folk song variants with melody-based features. In: Proceedings of ISMIR, Kobe, Japan, pp. 597–601 (2009)
41. Lilypond music notation software (2010). Retrieved from http://lilypond.org
42. Pikrakis, A., Giannakopoulos, T., Theodoridis, S.: A speech/music discriminator of radio recordings based on dynamic programming and bayesian networks. IEEE Trans. Multimed. **10**, 846–857 (2008)
43. Scheirer, E., Slaney, M.: Construction and evaluation of a robust multifeature speech/music discriminator. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1331–1334 (1997)
44. Tzanetakis, G., Cook, P.: Multifeature audio segmentation for browsing and annotation. In: Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 103–106 (1999)
45. Marolt, M.: Probabilistic segmentation and labeling of ethnomusicological field recordings. In: Proceedings of ISMIR, Kobe Japan (2009)
46. Kovačič, M.: New contexts, esthetics, and transfer in bell-chiming tradition. Slovene Stud. **29**, 19–34 (2007)
47. Klapuri, A., Davy, M. (eds.): Signal Processing Methods for Music Transcription. Springer, New York (2006)
48. Marolt, M.: Non-negative matrix factorization with selective sparsity constraints for transcription of bell chiming recordings. In: Proceedings of Sound and Music Conference, Porto, Portugal (2009)