Music Publishing

Alberto Manuel B. Simões
albie@alfarrabio.di.uminho.pt

José João Dias de Almeida
jj@di.uminho.pt

Projecto Natura
Departamento de Informática
Universidade do Minho
Braga - Portugal

Abstract

Current music publishing in the Internet is mainly concerned with sound publishing. We claim that music publishing is not only to make sound available but also to define relations between a set of music objects like music scores, guitar chords, lyrics and their meta-data. We want an easy way to publish music in the Internet, to make high quality paper booklets and even to create Audio CD's. In this document we present a workbench for music publishing based on open formats, using open-source tools and script programming over them. The workbench is based on an archive specification written in a text-based format which includes sound references, music scores, chords and lyrics and their meta-information.

Keywords: Music publishing, web, paper and CD publishing

1 Introduction

In the last decade Internet has been growing in a fantastic way. One of the main reasons for this explosion is the large amount of information people want to make available to other people.

While this proliferation of information is good, some areas of knowledge are being totally ignored in respect with Electronic Publishing. One example of these areas is Music.

While one could argue that a lot of people publish music (looking to the thousand of MIDI, MP3 or .au files in the web), what they are publishing is sound. Some other Internet sites present lots of lyrics but with lack of sound information. Some other, include some chord notations for guitar in the text, using HTML tags for formating, or using spaces and different lines to put them on the correct place.

This article will present our experience on music publishing, presenting the tools and file formats involved.

1.1 Architecture for a Music Archive

A music archive should contain various type of information, when available:

- <u>music scores</u>, to be printed and played by any musician, in a music score format or an image format — for music score format we can use abc or LilyPond text formats, presented on section 2.1;

- <u>guitar chords</u> — there is also open formats to define and use guitar chords (Chord or its extended version ChordPro) and a lot of tools to edit them. On this article we will use interchangeably Chord or ChordPro referring to the same format as a whole. These will be discussed on section 2.2;

- <u>music lyrics</u>, such that the user can sing them — these can be included directly on the `abc` or LilyPond notations as on the Chords notation; if we have the music on both formats (`abc` and Chords) the idea would be to replicate the minimum information (to associate the lyrics with the notes, in the case of `abc`, and to associate the chords to the correct position on the lyrics, in the case of Chords. Remaining verses could be putted together on another file, or together with some meta-information;

- <u>sound files</u> with the music being played — these files can be MIDI files generated from the music scores (there are tools to convert `abc` and LilyPond notations to MIDI, and vice-versa), or live recordings on `.wav` or `.mp3` format;

- <u>meta-information</u> like authors, players, composers, lyricists, etc; Meta-data gives us a natural way of cataloging musics but also a way to connect documents by their type, author, date, etc. For represent meta-information we are using a simple XML document, as there are lot of XML tools freely available.

When choosing the file formats for our archive projects we had in consideration:

- information should be stored for durability: this means we need to use open (specially ASCII) formats;

- files should be platform independent: the user should be able to use them on any operating system;

- tools to see, use and/or modify the information should be open-source, available to any user;

We want to publish music and related information on the web, but not only putting there the information. In fact, it should be a dynamic web-site with musics cataloged, search-able and browsable with an ontology driving the searches. The web publication should give access to all the information we have on the archive but also to some other kind of publishing: audio-CD and paper publishing.

2   Quick survey of Open Source Music Publishing tools

On this section we try to give a quick vision of some tools that exists freely available to typeset music scores, to typeset chords, to create musics, song books and so on.

Figure 1 shows some of the file types and tools presented on this section. It is shown here so that the reader can understand how the files can be converted.
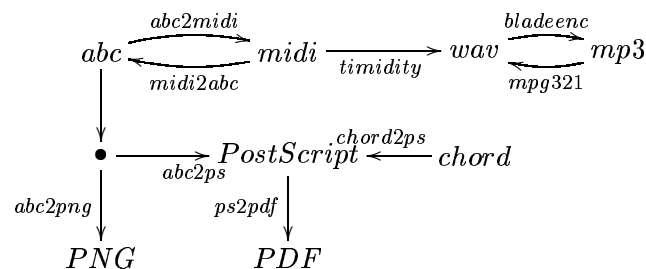

Figure 1: Graph of processes

2.1 Music Scores

For music scores we can use any of the file formats used by score editors. Main problems with these formats is information durability, platform independence and batch processing. Most graphical interfaces to compose music are very good, make it possible to export MIDI files, produce paper prints, export images. All this would be very nice if it could be done directly from a command line in order to cooperate with other tools on complex batch tasks, and if the software was free.

Only recently information durability has been taken care of. In fact, with the XML advent, anything using XML is announced as platform and application independent. This is true for XML but also for any other open syntax and, specially, when this syntax is written on plain text.

When we started collecting music scores XML was already being used but there was no XML based standard or proposal for music notation. Meanwhile, many score archives were found on the Internet using a notation named abc.

This abc notation is based on plain text, very compact and with many associated tools available on the Internet. abc is an intuitive language using only ASCII characters to represent music scores. It supports single or multiple staff musics, lyrics and guitar chords.

Figure 2 is an abc file content example and figure 3 the respective generated output using one of the many tools available on the Internet to process this notation.

```
X: 1
K: C
C D E F G A B c | c d e f g a b c' |
```

Figure 2: abc source file

Figure 3: Generated score output

While simple, this syntax is very powerful making it possible to produce professional looking scores. On the other hand, there are some graphical tools to typeset abc using a mouse, like tk-abc[5]. To learn more about abc you can read a nice tutorial written by Guido Gonzato[1].

But, not all are good reasons to use abc. On some cases, you have a sheet of paper with your score, and a recorded version on MP3, for example. Of course the MP3 sound version is better than any MIDI file, and to rewrite all the tune on abc need a lot of time. So, we think that a good PDF or PNG file for a scanned score and the associated MP3 is better than nothing, and as such, we consider this case on our workbench.

Other option, maintaining the open format idea, is the use of Lilypond[2], a GNU notation very similar to abc. On this notation, to produce the same output presented on figure 3, you need to write the code from figure 4.

While the notation is slightly more complex, it is a little more powerful than abc. If you think abc is really more simple, and can do whatever you want, use it. LilyPond distribution includes tools to generate LilyPond files from abc source files (abc2ly).

Both abc and LilyPond formats can include lyrics. Figure 5 shows our previous abc format with lyrics (the Latin name of the notes).

Both formats can be converted to MusiXTEX[7], a powerful LATEX like package to produce high quality music score Postscript pages. These can be included on any document we like.

Again, they can be converted to MIDI, too, which makes us able to listen to them.

```
\include "paper16.ly"
\score {
   \notes
   \sequential {
      \time 4/4;
      \key c;

   c,8 d, e, f, g, a, b, c | c d e f g a b c' |
   }
}
```

Figure 4: LilyPond source file

```
X: 1
K: C
C D E F G A B c | c d e f g a b c' |
W: do re mi fa sol la si do do re mi fa sol la si do
```

Figure 5: `abc` source file with Lyrics

2.2 Chords

To typeset lyrics including guitar chords there are various projects. For example GuitarTeX[4] uses a standard *chord* notation with some extensions, and a Graphical User Interface to help typesetting them.

Another interesting project is KGuitar[3], a graphical front-end for KDE with an editor for chords and tablature, with an integrated synthesizer.

The most common used notation is *Chord* or *ChordPro*. Follows a simple example:

```
{title:Lemon Tree}
{st:Will Holt}
When [D]I was [A7]just a [D]lad of ten, my father [A7]said to [D]me,
"Come here and [A7]take a [D]lesson from the [A7]lovely lemon [D]tree."
"Don't [G]put your [D7]faith in [G]love, my boy," my father [D7]said to [G]me,
"I fear you'll [D7]find that [G]love is like the lovely [D7]lemon [G]tree."

{c:Chorus:}
{soc}
    Lemon [D]tree, very pretty, and the lemon flower is [A7]sweet,
    But the fruit of the lemon is impossible to [D]eat.
    Lemon [D]tree, very pretty, and the lemon flower is [A7]sweet,
    But the fruit of the lemon is impossible to [D]eat.
{eoc}
```

Meta information is written between curly braces (`c:` − comment, `soc` − start of chorus, `eoc` − end of chorus). Remaining text is lyrics. Letters and numbers between `[...]` are chord notation which will be placed above the text when renderizing it.

3 Music Archives

This section will present some real examples of music archives on the Internet. The first two sections shows two examples of archives we are working on. Last section presents some other music archives available on the Internet.

Before starting describing our two examples we should explain why we have two projects about the same subject (Music Publishing) and what are their differences.
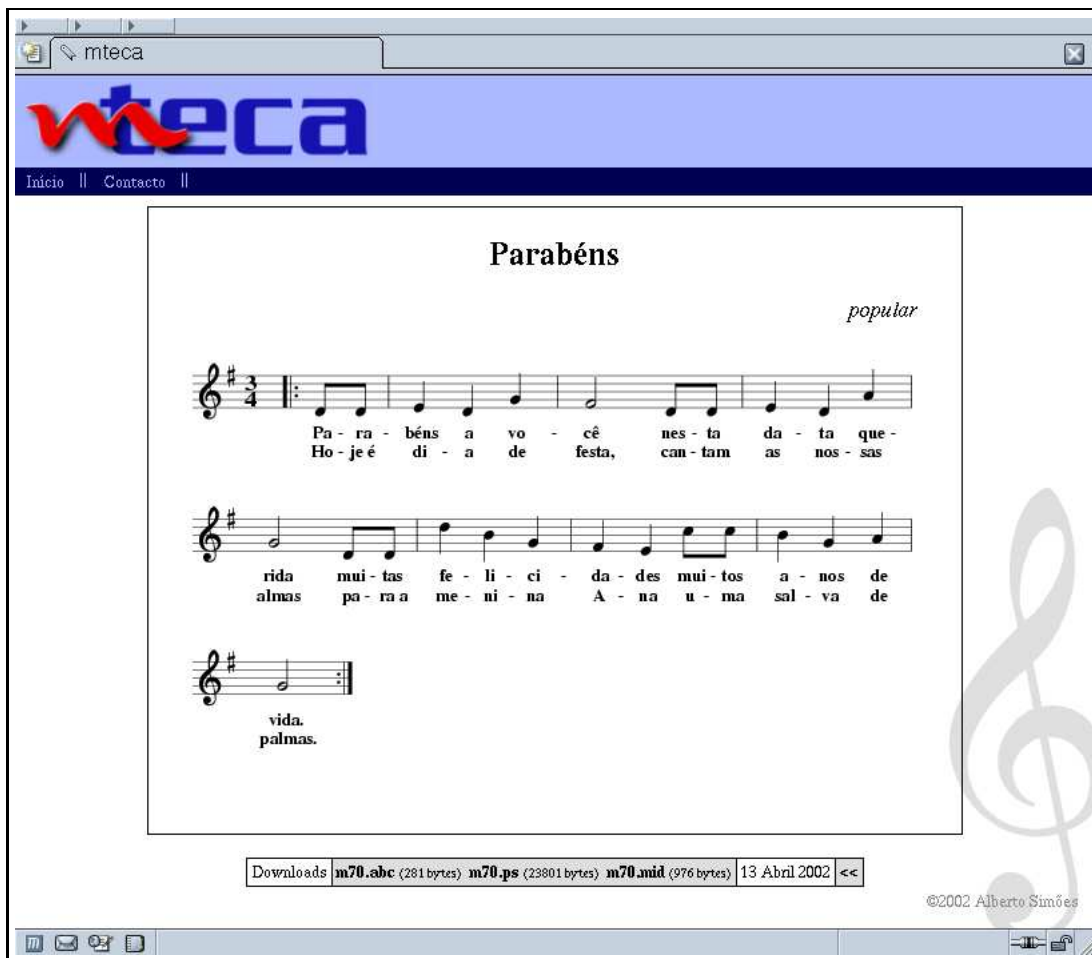
AMPP (Arquivo de Música Popular Portuguesa — Portuguese Popular Music Archive) is older than mTeca and grown as a hobby of collecting musics. As new musics were arriving with new formats, they were introduced in the archive. This means that AMPP is an heterogeneous archive, with music on many different formats: only the lyrics, lyrics with chords, lyrics with the music score, only the music score, and so on.

mTeca is based only on the `abc` format. While it is extensible to include other formats, the main goal is to contain music scores: so, chord format or lyrics only is not interesting for this project.

We are in the process of joining these two projects. Meanwhile, both of them are growing independently trying to get new ideas for the future common archive.

3.1 mTeca

This case study (`http://alfarrabio.di.uminho.pt/mteca`) intends to be a music library on the Internet. While the first idea was to compile any kind of music, at the moment it contains only liturgic music and an half dozen of popular music. Figure 6 shows a screen-shot for a document on this library.

Figure 6: mTeca screen-shot

This archive is based on `abc` providing the user with a web preview of the music score, access to the respective `abc` source file, to the postscript ready to print and a MIDI file to listen to.

The archive meta-information is stored on the `abc` files and on an XML catalog file, which

list all the library entries. This catalog is used together with an ontology to create a digital library[6].

While all the web-site is dynamic, music web-pages are static as the process of image files creation is slightly indirect, needing to convert first to Postscript and later to PNG images. On bigger music scores (more than one page) the conversion from Postscript creates a set of PNG images, for each of the pages. These files are cropped and the HTML embed all these files.

The abc publishing process can be described as:

$$catalog \rightarrow \begin{cases} abc \rightarrow \begin{cases} midi \\ postscript \rightarrow png \end{cases} \\ \text{HTML} \end{cases}$$

The files mTeca really archives are the catalog and the abc file for each one of the musics available. Everything else is automatically generated from them.

3.1.1 mTeca Future Work

At the moment mTeca has only a web existence. It is possible to download a MIDI file or Postscript file because they are available on the Internet. But we want something more: publication of audio CDs and song book publication:

- **Disc publication**
  For sound publication on mTeca we should use the created MIDI files. The files should be converted to raw audio formats as `wav` or `mp3`. This can be simply done using `TiMidity++`, which can be used to listen to MIDI files or to write the sound on `wav` format.

  But, to create an Audio CD we need something more: to compile the files on a ISO sound CD to be burned, and to create some kind of description about the CD contents. With this in mind we defined a XML format defined by the DTD presented on figure 7.

```
<!ELEMENT cd_toc (title, track+)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT track (title, author, file)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT file (#PCDATA)>
```

Figure 7: DTD for Disc Publication

  This format, while simple, contains all the information needed to compile a professional audio CD. It is possible to add times to the music list using a music player or one of the many available Perl modules to deal with sound files.

  All this is defined, but the tool is still under development.

- **Paper publication**
  As presented above, there are tools to convert all the formats to Postscript or PDF. This means that to publish this on paper should be simple.

  We want to publish the musics on a professional looking song book (something already possible with AMPP archive). As with the CD creation, we can define an XML meta-document with the structure of the book. This structure can be processed to create a LaTeX file with indexes, numbered pages and so on, ready to be printed or to make available on the Internet.

3.2 AMPP archive

AMPP is a digital library of Portuguese popular music. The basic contents of AMPP is songs (using a archive lyrics textual format `lyra`), and information about groups.

Each `lyra` document has:

- Meta information – information about the music as complete as possible (title, singer, author of the lyrics, author of the music, MP3, URL of the music, etc);

- Scores (in `abc` notation), if available;

- Lyrics;

- Chords (embedded in the lyrics; in a ChordPro like notation) if available;

- Textual notes (in HTML) if available.

Following we can see an example of a description of a music with `abc` score, lyrics, and Meta-data.

```
title: Canário
author: popular: Trás-os-Montes
comm: Constantim (Miranda); cantado pela Sra. Maria Cristal 1982
from: J.João
type: rimance

<abc>
M: 2/4
K: Dm
Q: 1/4=90
L: 1/8
A2|d2>d2| d2 ^ce | d2 Ac | c2>B2 | d2 c>B | A4 |
w: Es-ta ma-nhã fui à ca-ça lin-do ca-ná-rio ca-cei
z2 D2| A2>B2 | A2 GF | E2>F2 | G2 F2| E2 AG | F2 E2 | D4 |]
w: pa-ra tra-zer de pre-sen-te à fi-lha do* no-sso rei
</abc>

Esta manhã fui à caça
lindo canário cacei
para trazer de presente
à filha do nosso rei

A filha do nosso rei
ela era brasileira
mandou fazer uma gaiola
da mais fininha madeira
```

This textual definition, can be used to build several views, such as:

- HTML page of the music: build with the image of the score (abc2png:abc → image) and lyrics, and several links,

- MIDI and Karaoke sound files (using abc2midi),

- WAV and MP3 sound (abc → MIDI → WAV),

- PDF view of the music (lyrics → LaTeX; abc → PS to be imported by LaTeX with include graphics). This LaTeX file can be used in song books.

If the music had guitar chords, in order to build the views, we must build the images (PNG for HTML pages and EPS for LaTeX) of the chords used in the music definition.

3.2.1 From `lyra` to PDF

After processing the previous example with the command `lyr2any -tex file` we obtain:

Figure 8: PDF view of the song *Canário*

```
title: O sol perguntou à lua
author: popular: Açores
from:Tó Campos
chorddef: D#dim xx1212

O [Gm]Sol [Gm7]perguntou à [D#dim]Lua [D7]
O Sol perguntou à Lua
Quando'[Gm]a, quando ha[Gm7]vera amanha[D#dim]cer [D7]
Quando'a, quando havera amanhecer

À vista dos olhos teus
À vista dos olhos teus
Que vem, que vem o Sol cá fazer
Que vem, que vem o Sol cá fazer

[D7]E o Sol préguntou à [Gm]Lua
quando ha[Gm7]vera amanhe[D#dim]cer [D7]
```

In this example we can see the use of chords. In order to make a PDF view of this song we need the definition of each chord. The command `lyr2any` has an internal chord database that

contains the most common chords. All of the chords used in this song are known in this chord base. We included the definition of `D#dim` in order to show how the user can define new chords.

In order to render chords, `lyr2any` extracts the list of chords used in the song and builds the PNG and the EPS images of the chords.

Figure 9: PDF view of a `lyra` file with chords

3.2.2 From `lyra` to HTML archive

In order to make an archive, it is necessary to add connections between similar documents.

In AMPP we add a general view of the archive based on:

- author, composer, lyricists, – connecting authors to the list of the songs they wrote (based on the meta-data present in the lyra files);

- type of song – connecting types to the songs;

- date – linking songs by date order.

In order to improve author view, it is provided a way to define:

- which musicians belong to each band;

- external pages connected with an author;

- textual information about the authors.

From this definition all the indexes and reverse links are generated.

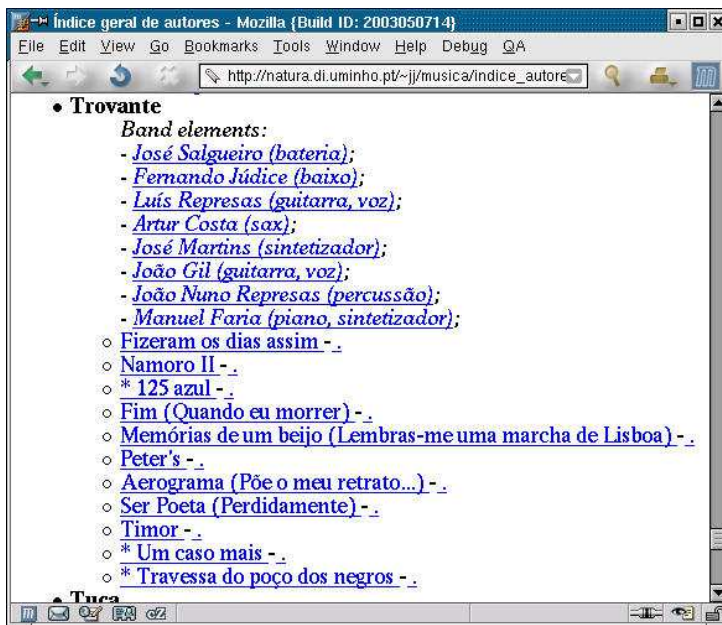In figure 10, we can see an example of a detail of this kind of indexes.

Figure 10: Author index: connections of the group *Trovante*

3.2.3 From `lyra` to song book

It is very important to be able to build song books. The song book generator tool is using the translation (lyra → LATEX) applied to a set of lyra documents.

The song book generator:

- makes table of contents, conceptual index,

- can incorporate LATEX text,

- can produce A4, A5, booklets.

In AMPP the we have near 900 songs archived. The song book generator can use pattern queries (typically over meta-data) to define the set of songs to be included in the book.

3.3 Other Projects on the Internet

On this subsection we will present a list of available projects on the Internet doing music publishing.

- **Mutopia** — http://www.mutopiaproject.org/
  Mutopia is an LilyPond based music score archive. All the provided files are free to download, print out, perform and distribute. At the moment of writing, Mutopia has 283 musics for download from Beethoven to Joplin, from Harp to Piano or full Orchestra.

- **Classical Music Archives** — http://www.classicalarchives.com/
  The Classical Music Archives is a huge collection of MIDI files for classical music. It contains also some MP3 files of human performers to be downloaded and listen to. This project limits the number of downloads unless you subscribe to the archives;

- **The Sheet Music Archive** — http://www.sheetmusicarchive.net/
  This is an archive of music scores on PDF format. These files are old books digitalizations where copyright as expired. It contains a lot of files available for download;

- **Homepage of La Folia** — http://members.chello.nl/folia/
  As the author says *"for those people who are interested in the folia theme there is a disturbing lack of good sources"*. This is a really interesting archive of music about the folia theme.

- **The On-Line Guitar Archive** — http://www.olga.net/
  OLGA is a library of files that shows how to play songs on guitar. It is an archive of tablature files.

4   Conclusions and Future Work

To publish music on the Internet is not a difficult task. In fact, there are simple tools available which does most of the work.

One of the biggest problems to music publishing is that here is no good standard for music typesetting. MusicXML is too much verbose, as MusicTeX. Lilypond, on the other hand, is very similar to abc and could be a nice choice. Having one standard, commercial tools will start export options for this formats, as would open-source community create tools to use this file format. Basically, this area of knowledge is needing something like MathML has done to mathematical community.

References

[1] Guido Gonzato. *Typesetting Music with ABC*, version 1.0.3pre1 edition, March 2003.

[2] Jan Nieuwnhuizen Han-Wen Nienhuys and Adrian Mariano. *LilyPond - The GNU Project Typesetter*, 1999.

[3] Kguitar: a tabulature editing gui with midi synthesizers support. http://kguitar. sourceforge.net/.

[4] Joachim Miltz. *GuitarTeX - a tool for guitarists print and play song sheets from their chord or chordpro files*, 2003.

[5] Jean-François Moine. A graphical user interface for music edition. http://moinejf.free. fr/tclabc.html.

[6] Alberto Manuel Simões and José João Almeida. Library::* — a toolkit for digital libraries. In *ElPub 2002 - Technology Interactions*, 2002.

[7] Daniel Taupin. *MusiXTeX— Using TeX to write polyphonic or instrumental music*. Laboratoire de Physique des Solides, version t.111 edition, April 2003.