

Sometime in 2015, Riyaz Shaikh on thesession.org <https://thesession.org/discussions/38035> posed a question about tools to convert sheet music to abc notation. Riyaz had an interesting problem. He happened to stumble upon a tome (so to speak) of Scottish tunes. They were apparently transcribed by a traveling fiddler from Scotland, and shared with the Learning Session group in San Antonio, TX. With over 1000 tunes in the tome, Riyaz was understandably reluctant to hand-code them in abc notation - so he polled the forum on thesession.org.

That opened a whole can of worms across a number of interesting threads as the project took shape. Just for starters, there were issues with titles. There's an Irish saying: "There are two sides to every story and 12 names to every tune." Truer words were never spoken. The tome, which we came to call "Tom Buchanan's Scottish Tome", suffers from a series of sequential copying machine ventures and not all of it is too clear. The transcriber - whom we assume to be Tom Buchanan - did a magnificent job of writing down tunes on staff-lined paper. With a few caveats that is. When there was insufficient space at the top of a tune for a title, it might get placed at the bottom. Or the middle. Sometimes the multiple passes through copying machines truncated the top of the page and there was no apparent title at all. Great fun. Spelling was also an issue. "Stone Feild" is almost certainly supposed to be "Stone Field" but perhaps not. If the transcription doesn't resemble a tune named "Stone Field" found somewhere else, it might be an entirely new tune. Comparisons have to be made. There were, and still are, issues with the types of tunes in the tome. They are not in any kind of order associated with the more common tune books found on the web. There is some effort to place tunes in categories such as reels or jigs and such but not so much that it's readily apparent when thumbing through the 652 pages Tom transcribed. Occasionally, there might be a word in the margin defining the section as "Reels" but more often not. I'm not going to fault the gentleman, it must have been a Herculean effort to transcribe all of the material in the tome. It is just a bit haphazard in some respects.

There were three of us in the project, Riyaz Shaikh, Jo Wegner and me - Carl Allison. We formed at very informal working group. Riyaz provided scans of the pages and Jo and I took to creating the abc notation. There was more than a bit of discussion on thesession.org as to what format the output should be in. Since the underlying tune repository of thesession.org is abc notation we decided that abc notation would be the format chosen. No doubt there are other music notation formats that are much more capable but they are not "open source" and sometimes they are quite expensive. ABC notation is open, free and has multiple applications available across the three major user computing environments (although linux has very few...). Besides, the Tome is just Irish and Scottish tunes - right? Simple stuff, abc notation should be entirely capable of the job. Or so I thought. If one were to drive through the many posts I made on thesession forum requesting help for this symbol and that symbol, it would become obvious that abc notation does not cover every possible embellishment that might be found in a tune. Jo opted to hand write the 50 or so tunes she had time available for and I opted to try and automate the process using software. It took us just over 10 months to convert the 1102 tunes to abc notation. Before anyone becomes vocal about how bad - or good, the effort was, I have to mention that no one is infallible. There are mistakes. In the sense that this is an effort with open source software and standards, so too are the tunes "open source" in the sense that they can be corrected. Sometimes there are things in the original pdf scans that are simply not available in abc notation such as straight-line slurs and a few other ornamentations.

A caveat about copyright. Undoubtedly, some of the material is covered by some form of copyright. It was never the intention to skirt around copyrighted material but simply put - I didn't want to spend an extra month or two - or more - trying to chase down each tune. If the Tome runs afoul of someone, they need to contact me so that I can remove the offending material.

Scans vs ABC notation only:

Ah yes. The scans. Anytime someone brings up abc notation on the session's forum, there are vocal elements that eschew the underlying technology of abc notation because it is felt that the true Irish Trad musician never uses sheet notation (more commonly referred to as "dots") and therefore, there is no need for abc notation. Equally vocal are the comments that abc notation is difficult to read and bears little resemblance to music. To that, I would agree. I will state though, that abc notation provides a very capable mechanism to generate sheet scores (dots) that are very legible and easy to read and at the same time, because abc notation is text, it is very, very efficient in terms of computer storage space. Not such a bad thing. I don't personally try to read abc notation and directly play it on an instrument but abc notation has allowed me to learn to read music scores enough to memorize the tunes and get me to the "normal" mode of playing Irish Traditional Music (ITM). I've digressed a bit here though so back to the scans. There were vocal elements on the forum that scans were important and that to merely provide abc notation would be an incomplete solution. In the long run, the working group decided that in addition to the abc notation, the scans would be provided as well. There were further discussion on the merits of resolution in pdf files vs disk space and presumably, download speeds as well. We tried to achieve the best balance between optimum and reasonable quality.

The tome, all 652 pages of it, had one or more tunes per page and sometimes a tune that would be continued on a following page or perhaps be scored across multiple pages. In an effort to keep tunes vs pages a somewhat coherent mess, the tunes are numbered by their start page number, a decimal and a position on the page such as "1" if it is the first tune. "4" if it's the fourth and so forth. If the tune is numbered "xxx.0", the zero means it's the only tune on the page. When someone looks at the abc notation or the pdf output of the abc editor, the number allows the viewer to refer back to the correct original scan. It may not be entirely clever but it does work.

Automating the conversion process: I use a Macintosh so the software I use might be unique to that platform. I'm fairly certain that applications with the same capabilities exist on Windows and Linux. I used EasyABC which is cross-platform. In addition to EasyABC, I used QuicKeys, the unix terminal running shell-scripts (which I wrote), Photoshop Elements, GraphicConverter, Apple's Preview and Finale PrintMusic's SmartScore Lite module. Indexes, table of contents or permuted indexes were created with a combination of unix terminal commands and MicroSoft Word.

Here's how that process goes. Somewhat.

1. Convert pdf files to a tiff files. GraphicConverter in batch conversion mode. Output to grayscale, 600 dpi.
2. Manual separation of each tune into a separate graphic file. Adobe Photoshop Elements. Before the individual tunes are separated, the entire page has the background subtracted, manually rotated to the correct orientation, staff lines re-drawn manually if missing and if necessary, staff symbols copied to lines where they're missing due to copying machine truncation. Then tune tunes get separated. Rather than a quadrillion or so key-strokes, a lot of the steps that would normally require manual interaction were replaced by QuicKeys macros.
3. Load the file into the SmartScore Lite module using a QuicKeys macro and start the processing and then save the output (if there is any, hand-written scores are difficult for computers to discern). The output is saved as a .xml file. Another QuicKeys macro opens that in EasyABC.

4. EasyABC's code field which is text and that's where you operate on the code. The xml stuff from PrintMusic though, has essentially the wrong line ending character so a QuicKeys macro copies the entire text field, runs it through a unix shell script and then takes that output to replace the original code and saves the file while giving the file the correct name from a master name and index number file. At that point, it's necessary for the human element to get involved with adding accents, fixing errors, adding chords and a plethora of other actions. One of the worst has to be doing triplet slurs so a QuicKeys macro does the heavy lifting on that action as well. There were a lot of QuicKeys macros towards the end and their utility saved me a tremendous amount of labor plus they're consistent whereas I'm not always that consistent.

5. Listen to the midi output from EasyABC and make corrections as necessary. Fix timing errors, repeats that fail and so on.

There are a bunch of other processing details which have sort of fallen out of my brain but they were mostly shell scripts to convert accent marks etc through an entire tune rather than me doing it by hand.

There are a lot compromises in the abc notation in that abc notation itself doesn't support a lot of ornamentation found in the original scans:

Straight line slurs don't exist, regular slurs are substituted instead.

Sometimes repeats were added at the beginning of a "B" section if EasyABC failed to repeat. That one isn't 100% consistent.

Sometimes the key wasn't called out. Referring to the tune on thesession.org and other abc notation sites generally gave the correct key. If not, experimenting provided the correct sound. Hopefully.

There were a lot of weirdnesses, enough that I don't recall them all.

Now that the project is "finished", I'm waiting for feedback on errors that get spotted by someone else and not me. When I hear of them, I'll fix them.

Carl Allison 2022

bouzoukidude2021 <at> gmail <dot> com