

How AI Made a Purely Music Site Profitable

I started the Classical Archives (classicalarchives.com) in 1994, as a hobby while living in Hong Kong. I had moved there (to license some of my IP) from New York where I used to be quite involved in the classical music scene. Many of my friends were either composers or performers and I used to regularly attend concerts and recitals (and organized some myself.) When I arrived in Hong Kong, I realized that the main culture there was (1) eating and (2) making money. There was scant classical music.

1994 was just a couple of years after the beginning of the web. It was virgin territory. At that time, I used to create and collect classical music MIDI files. MIDI (Musical Instrument Digital Interface) was a terrific way to store and manipulate scores that one could perform through a synthesizer (which is now part of all computer audio cards.)

I started Classical Archives by offering free access to my own files – all of which were from composers long in the Public Domain. I took great pains in categorizing everything accurately. And at the bottom of my pages, I invited visitors to submit their own files to add to the collection – with the only promise that I would continue to take pride in categorizing each entry properly. And before long, I was hosting the largest collection of classical MIDI files in the world.

Then, in 1999, after returning to the States (I had moved to Palo Alto to be close to Stanford and to benefit from the ferment in Silicon Valley), I received an offer to sell the Archives. As I was not keen on selling, the investors and I agreed that they would provide some capital so we could add live recordings to the site, again, making sure that all items would be well categorized.

I recall the horror expressed by the major labels when I approached them then to license and place their tracks on our site! (This was the height of the Napster controversy.) So we elected to find live recordings from artists who were more trusting of our ethos. I engaged a wonderfully well-informed and well connected person in Moscow with the brief to go a concert every day and approach the artists in the Green Room to propose: “You have world-class talent but you are fairly unknown in the West. Let us put your work on our site. We won’t charge you anything.” Our “Ambassador” went to work and before long we had tens of thousands of high-quality recordings on the Archives.

The record labels then came around. We signed hundreds of labels and started to ingest close to a million tracks. This is when we realized that if we wanted to maintain the very high quality of our composer/work/movement/artist/instrument/period classification, we would need a dozen highly trained full-time musicologists on staff.

At that time, if you searched for “Beethoven” on iTunes, you’d get back a movie about a dog, some funk rock bands from Scandinavia, and a sprinkle of actual Beethoven works—performances of the most accessible repertoire.

So I went to work to develop a technological solution to this issue. I convinced the creator of the Music Genome Project (a Stanford musicology PhD who had just left Pandora) to join me. We created a canonical database of classical music compositions and performers and worked on an intelligent algorithm that allowed us to ingest and clean the fairly dirty metadata coming from the labels. We called this the Contextual Metadata Engine (CME).

The CME now allows the Classical Archives to ingest extremely efficiently all new material while maintaining the high quality of the site's organization. Instead of a small army of musicologists, we only have one on staff. The quality of the site, the loyalty of our users and the reduction in costs allowed us to become cash-flow positive years ago and we have been profitable ever since.