

Chapter 1

Artist Resolution

Let us start with the simple assertion “An artist is defined by what I want it to be”

After examining hundreds of thousands of mp3s we find that artists are defined by their name.

As there is not a standards agency that doles out band identifiers. We must derive one.

I start with the assertion that each “band” must be assigned an immutable arbitrarily assigned identifier.

- Beatles
- The Beatles
- Beatles, The

What is the “Name by which you would like to have The Beatles identified?”

Simple is good.

Let us define a table and call it artist.

The unique identifier for artist shall be “artist_id”.

This identifier should be immutable, that is should never change.

A property of artist is the preferred name for the artist.

Natural keys.

artist_id	artist identifier	immutable means of specifying an artist
artist_name	The preferred user friendly representation of an artist’s name	

The artist will be referred to in relational objects by the identifier.

Now we need some way of assigning an artist_id to various representations of artists names.

todo with various singers?

```
select upper(artist_name), count(*)
from mp3
where upper(artist_name) like '%BEATLES%'
group by upper(artist_name)
```

BEATLES, THE	28
BEATLES	81
THE BEATLES	2
BEATLESLES	1
THE BEATLES	559

Artist Name	Count
Beatles, The	28
Beatles	81
The Beatles	2
Beatlesles	1
The Beatles	559

Popularity is a good indicator of preference. Locating
notes
inspection of data