

Christy Crawl, Founder, ProMusicDB Presentation notes

Authority Control Interest Group - ALA Annual Conference

Las Vegas, NV, June 29<sup>th</sup>, 2014



### **SLIDE 1: Introduction**

Hello everyone, it is good to be with you again. My name is Christy Crawl, and I'm the Founder and Project Director of ProMusicDB: the Professional Music Credits Database. Before diving into the recent developments of the ProMusicDB ID and our Working Paper, I'd like to summarize the project's scope, our path, and our initiatives for a few minutes for those of you whom we might be new to.

First, our "position" as we embark on this endeavor and as we ask for contribution, feedback, collaboration, and assistance from the academic community, the library community, professional music organizations and their members, and the National Endowments.

### **SLIDE 2: Music Metadata today = MUSIC HISTORY TOMORROW**

Our musical heritage has slammed into the digital age. It is eroding because of the impact. Yes, we all should be concerned with the preservation of the United States' historical musical collections and the ongoing referencing of born digital works. But unless there is a joint effort between music-related "authoritative" agencies to help codify a music credit standard and help establish a referencing framework in which Music Librarians, Professional Music Organizations, and Music Distributors can willingly participate, individual endeavors will be wasting money and time.

In essence, the music business is now a music metadata business. The metadata is in “the cloud,” or more simply, on the ground in dataservers. No platform yet pulls it all together, organizes it, and efficiently and elegantly distributes it in the form it is needed to its individual queries. The future of the digital age will demand that all Music-related Metadata work in harmony, and not in discord, with other pre-existing and yet to be created Metadata. It cannot do this without a correct music credit standard. If we, as a culture, do not get this supporting music credit framework correct and in place soon, we will not be able to catch up.

We propose that ProMusicDB platform is positioned to supply the needed framework.

### **SLIDE 3: Where have we been prior to the ALA Annual Conference?**

Over the past year, we have engaged the Professional Music Unions, the Music Library Association, and the Entertainment Identifier Association as supporters and consentors in our mission. We have been on the “conference trail” introducing our endeavor and our quest for the utopian music credit standard system of authentication and verification across the country,

### **SLIDE 4**

for through our participation in metadata working groups across the academic, professional music organization, and record industry, we have concluded that the creation of ProMusicDB will serve many needs in the creation, monetization, and archiving of musical works and professional musical artists. I’ve relayed much of the information we have gained and much of our process in previous presentations, but wanted to share again with you our “Smithsonian of Music Credits” design, where you can see how all contributors “fit” and receive benefit.

This presentation, as our others, will be available on our website [promusicdb.org](http://promusicdb.org)

If I can focus your attention towards the bottom of the slide in the center, you will see our “Unique Identifier” placed, and that is what I am here to discuss in detail with you today -

### **SLIDE 5**

Along with sharing some tools we are developing, and what has been brought to us to solve by the American Federation of Musicians.

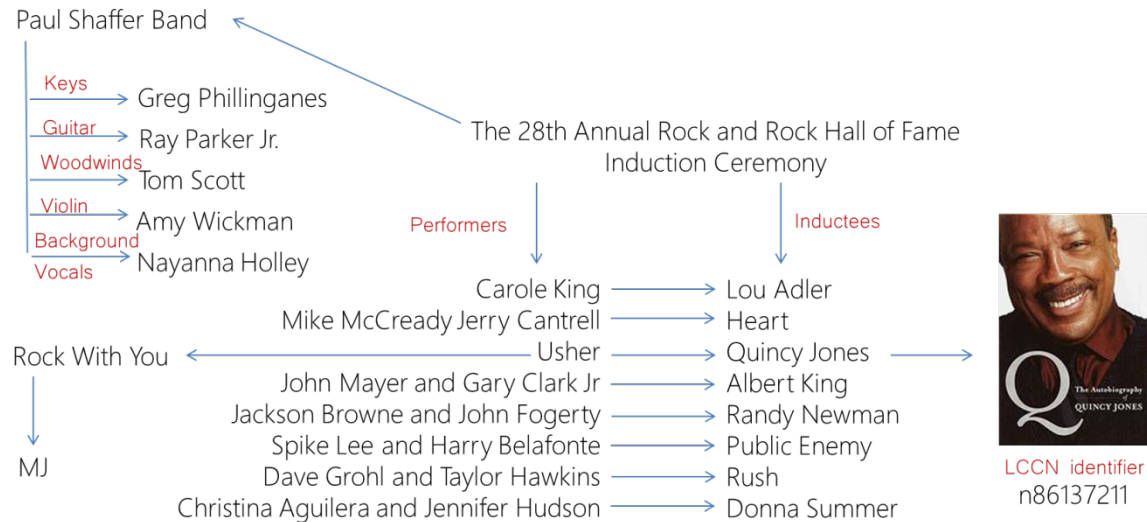
### **SLIDE 6: Publicly available online resources**

Music is increasingly being distributed, consumed and rated online. Various aspects of professional music and related activities are handled using online portals with holding databases of millions of records, artists, users, playlists, and ratings and other activities. However these systems are musical artifact centric and hence focusing more on published work. This is increasingly becoming a challenge as more resources are becoming digital born. Professional music activity goes much beyond online consumption of music.

### **SLIDE 7: Industry only metadata resources**

Membership to music unions, professional rights organizations, awards and recognitions, live performances, and other professional services are mostly contained in organizational boundaries and not made available to public at large. In spite of being published on numerous music credit databases, music sharing portals, social networks, and online stores, music professionals cannot establish their credibility due to issues like missing information, duplicates, lack of accuracy and authenticity among other numerous limitations.

### **SLIDE 8: ProMusicDB ID Generation for Complexities in Musical Events and Performers**



**Figure 1:** Complexities in Professional Music Domain

This slide with music credit metadata from the Live but recorded for broadcast HBO Special Rock N Roll Hall of Fame Concert in April 2013 depicts complexities in professional music events. The usual practice of the recorded version of such events when published only contains major credits for lead performers, directors, composers, etc. Details of other music professionals and their contributions therefore remain invisible to general public. From the professional musician point of view however, professional association at such a prestigious event can be critical to all concerned music professionals. ProMusicDB therefore attempts to facilitate capturing, handling and furnishing such details by partnering with appropriate stakeholder organizations. Professional details hidden behind organizational boundaries captured in form of official documents, contracts and agreements, as well as session sheets can be captured and made available in the form of an enriched professional profile.

## SLIDE 9: ProMusicDB ID Cross References Search Engines

However, in order to be successful, the ProMusicDB effort will have to integrate well with all existing stakeholder organizations and their datasets. The most common way for music related queries to be made are through the most popular search engines. Google knowledge graph (Freebase<sup>1</sup>) is a community curated database of people, places and things containing the highest number of topics (~28M) and facts (195M) that are related to music. Additionally through the Schema.org initiative, major search providers are helping bring structure to facilitate better searching. Standards and practices from Semantic Web and Linked Open Data communities are making it possible to take integration to the next level.

## SLIDE 10: ProMusicDB ID Cross References Linked Open Data

<sup>1</sup> Freebase: <https://www.freebase.com/>

While these efforts have resulted in massive interlinked datasets, the accuracy and coverage have been challenged restricting their usage in the professional context. This issue is addressed by information science community through efforts in library catalogs; archival and cultural heritage curation. Authorities in form of **People, Places, Topics** and **Organizations** are appropriately linked with **creative works** and their **instances**, further enriched with **annotations** related to library holdings. Librarians, along with information experts, through extensive research help maintain curated records.

### **SLIDE 11: ProMusicDB ID - many IDs issued by different stakeholders**

For instance, music library community through rigorous standardization and massive curation efforts (BIBFRAME<sup>2</sup>, VIAF<sup>3</sup>) holds extensive data for published music records that are accessible to general public via services like WorldCat. Focusing on music domain, there are several efforts focusing on collecting, processing, and managing music-related information focusing on specific aspects like music credits (MusicBrainz, AllMusic/Rovi, Discogs), film & TV (EIDR), music analysis (EchoNest). Digital born resources and their various forms of activities on web are increasingly becoming critical part of music experience. Social content on various online services (Youtube, SoundCloud, etc.), music discovery and playlist services (Lastfm, Pandora, Spotify, Songza, etc.), Online stores (iTunes, Google Play, Amazon, etc.), and social networks (Facebook, MySpace, Twitter, etc.) are increasingly offering critical context that must be considered quite relevant to music professionals.

In summary, it becomes imperative for ProMusicDB to integrate and interoperate with all these existing datasets which are already used very widely in different context. In doing so, harmonization of various standards, deduplication, and other integration issues will have to be resolved.

### **SLIDE 12: ProMusicDB ID - RRHOF Stakeholders & IDs**

#### **So, there is a need for Unique ID for Professional Musicians and Related Professional Musical Events**

In order to be successful in this feat, it will required to capture and maintain various details for all individual music professionals, groups, organizations, creative works, professional projects, releases, events, and similar other professionally relevant aspects. It therefore becomes imperative that in the resulting database, all these entities should be accurately identified in a globally consistent manner and complex relationships among them should be discovered and represented such that scenario events such as the Rock N Roll Hall of Fame Concert 2013 can be accurately described and made available to stakeholders and the public.

### **SLIDE 13: ProMusicDB ID - Key Requirements**

#### **promusicdb ID: ID Generation Logic**

---

<sup>2</sup> BIBFRAME Model: <http://bibframe.org>

<sup>3</sup> Virtual International Authority File (VIAF): <http://viaf.org>

Our purpose in the creation of ProMusicDB is (a) to support disambiguation, (b) to accurately capture professional music activities, transactions, events etc. and (c) to provide trusted and verified information. In order to facilitate this requirement, we offer the ProMusicDB ID - an identifier designed to cover four aspects namely (i) structure (ii) information content (iii) encoding/mapping, and (iv) verification.

**Structure:** The structure of the ProMusicDB ID addresses the question of “What goes in to the ID?” Carefully selected data points that are useful in disambiguation at the same time should be able to accurately describe various kinds of entity in music domain. For this, the structure of the ID consists of a type identifier, a date field, a time field, a genre field, a credits/type and finally a system generated unique number.

This means, in order to generate the ID for any entity, all such data points should be known for that specific entity. To address this requirement, the ProMusicDB ID Generation process also facilitates a registration process for entities that do not have such information at the time of registration. For the individual music professional, ProMusicDB can act as a registration agent for artist not listed on ISNI. Similarly, for digital born resources ProMusicDB can act as Registration agent for listing on EIDR for the digital Film and TV product supply chain.

**Information Content:** The information content aspect addresses the question of “Where the data will come from?” ProMusicDB ID generation process can be triggered from various use case scenarios. For a given artist, a profile can be bootstrapped from existing music data sources listed in earlier. Data populated in this manner will be treated as unverified. Another use case can be curation by experts in library community. They can help establish relationships among artist and creative work by selecting appropriate classification schemes. Most music data portals provide capability for programmatic access to dataset through APIs. To aid this process of getting frequent updates from sources that are not exposed through API, advanced data extraction capabilities (through extractors, crawlers, connectors) offered by Kimono Labs<sup>4</sup>, and Import.io<sup>5</sup> can be utilized.

**Encoding and Mapping:** The encoding and mapping aspect addresses the question of “What should be the format?” Data extracted/retrieved from various partnering organizations can be in different formats and encoding. For consistency, they must be transformed while being extracted and integrated with ProMusicDB. For instance, for representing Genres, NARM repertoire codes will be used. For capturing credits, allowed value sets for artist roles specified in DDEX standard will be utilized.

**Verification:** The verification aspect addresses the question of “Where the supplied information is can be verified to be correct?” Data once captured and encoded will still be marked as unverified data in ProMusicDB system. Such records will be subject to verification workflow that will involve the partnering professional music organization. For instance, in case an artist is claiming composer credit for a specific work, it will be verified against official records of concerned music union/PRO or recording studio.

---

<sup>4</sup> Kimono Labs <http://www.kimonolabs.com/>

<sup>5</sup> Import.IO <http://import.io/>

## SLIDE 14: ProMusicDB ID - Components – Type Identifier

### Components of promusicdb ID

**Type Identifier:** Type identifier component in ProMusicDB will specify the type of the entity being identified. For music professionals the type identifier will be number 1. Number 5 is reserved for various types of events including live performances, tours, award ceremony, games, inauguration functions etc. The type identifier will determine the values in trailing components.

Value	Description	Example
1	Music Professional	Quincy Delight Jones Jr.
2	Professional Entity	La Phil
3	Musical Work	Rock with you
4	Musical Publication	Rock and Roll Hall of Fame + Museum live (DVD)
5	Live Event	2013 Rock and Roll Hall of Fame Induction Ceremony
6	Project	Recording Session

## SLIDE 15: ProMusicDB ID - Components – Date

**Date Component:** Next eight digits (place 2-9 of the id) are reserved to contain date component of the identifier. In case of ID for the music professional, the date component will contain information about the birthdate or date of joining of professional organization. In case of event this component will contain start date of the event. This will follow the MM-DD-YYYY convention for representing the date format. This data will be verified against records of music unions/PROs or any other partner organization offering verification service to ProMusicDB

## SLIDE 16: ProMusicDB ID - Components – Location



**Figure 1:** Los Angeles instance in Geonames database

**Location Component:** Following the date component, next seven places (place 10-16 of the id) are assigned to contain location information. In case of music professional, this field will provide information about birth place, or place of registration. In case of event ID, the location component will correspond to event venue. Source of the information can be either supplied by music professional or their representatives, or it can be from secondary information sources like websites, encyclopedia or other online sources. Supplied location information will be looked up on geonames<sup>6</sup> database and will be encoded as unique 7 digit Geonames codes. The location information will be verified against membership data maintained by partnering music unions, PRO or similar organizations where the professional has registered as a member.

## **SLIDE 17: ProMusicDB ID - Components – Genre and Credit**

Code	Description
01	Classical
02	Jazz
03	New Age
04	Adult Contemporary
05	Rock (Soft/Hard)
06	Country

**Genre Component:** Next two digits (place 17-18 of the id) are allocated to specify the genre. For music professionals, it will be genre of the first work being registered or in case of an established professional, the most contributed genre. For events, it will be the most common genre of the musical work being performed during the event. Genre information in ProMusicDB ID will be encoded in the form of standard lists of genre identified by NARM in the of NARM repertoire Codes. For a given work, the genre information supplied for the music professional or event at the time of registration in ProMusicDB database will be verified against the registration information available at relevant professional organizations. For music professionals, such information can be verified by music unions, PRO or similar organizations where the professional is registered as a member.

**Credit Component:** Following two digits (place 19-20 of the id) are allocated to hold credit/type information. For music professionals, these two digits will contain information on first or most contributed role. For standard list of valid credit values, DDEX artist role will be used. DDEX is a digital supply chain standard enjoying prominent compliance by key stakeholders in the music industry. DDEX standards provide allowed value set for possible artist roles include specialties

<sup>6</sup> Geonames Database (<http://www.geonames.org>)



like choreographer, designer, graphic artist, and ensemble among other general roles of director, composer and lead artists. By adapting this comprehensive value set, the ID becomes much more powerful in terms of specificity and interoperability that drives disambiguation.

For events, this will correspond to the kind of event. Events can be of various types including award ceremony, live performance, inauguration ceremony, academic events etc.

## **SLIDE 18: ProMusicDB ID - Components – Unique Number**

### **Unique Number Component**

Trailing 4 digits (places 21-24) will be assigned by the system to insure uniqueness of ID. This will be a system determined auto increment number that will be assigned by the system. Same logic will apply for all types of IDs. This being a system generated ID, no verification will be needed. Also, the encoding and other overheads will not occur in such assignment.

## **SLIDE 19: ProMusicDB ID - Examples**

Using our generator Logic, here are some examples for person Quincy Jones, an entity such as the LA Philharmonic, and the Event “Rock N Roll Hall of Fame Concert 2013”.

## **SLIDE 20: ProMusicDB ID -Bringing all IDs under our umbrella**

## **SLIDE 21: How ProMusicIDs generated in accordance with type**

## **SLIDE 22: Quincy profile example: NOW AVAILABLE TO VIEW on website**

## **SLIDE 23-24: ProMusicDB IDs generated for Slide 21 with our RRHOF example**

## **SLIDE 25: Nest steps, tools, and use-case**

