

Cooperation with MIC

Daniel Antal, CFA

2022-10-07

In 2022, most of the world's music is sold (licensed) via the automated systems of services like Spotify, YouTube, Apple, Deezer. Artists or their representatives get micropayments after uses of their work. Most music is introduced to music lovers via autonomous (AI-driven) recommender systems, such as Spotify's algorithmic playlists (Discover Weekly, etc.). Festival organizers and music clubs increasingly decide on whom to invite to their stages based on streaming services and social media metrics.

Most of the European musicians, particularly in less affluent EU states like in Lithuania, are self-released. They do not have a large music publisher or music label who manages their data, tests if the algorithms promote them well and if they appear on the correct playlists. These artists either try to work with the world's most complex data systems on their own, or they rely on the work of freelance promoters or very small labels that do not employ a permanent data engineer or IT professional. Listen Local aims to help them.

In our project, we aim to do three things: - We want to observe how AI algorithms learn and promote the music of a limited number of Lithuanian artists (old and young, male and female, solo and group.) - We make minor adjustments, for example, corrections of errors/omissions on Dbpedia and name authorities, and measure the effectiveness of our interventions (increased number of plays, listeners.) - Pilot a small, semantic web application that helps organizations like the Lithuanian Music Information Center, the National Library, or LATGA to better support artists' visibility and increase their potential to earn more royalties and be invited to new live performance stages or festivals.

There is a lot more that we could do if we could secure additional resources and funding, but with our current MusicAIRE grant, we are aiming for something small and tangible.

Observe and Research

We are creating several datasets that investigate the data quality of some Lithuanian artists' digital presence and establish the value and priority of correct data. As a first step, we observe a few dozen artists who are featured on the Lithuanian Music Information Centre Database in the Folk genre. We will assess the data quality of these artists in several web services. We chose Folk because it is more popular than Contemporary Classical music and less popular than Pop/Rock.

Our discoveries

- We pseudo-randomly picked artists and check their data quality on Wikidata, the Library of Congress, GND, Spotify, YouTube, MusicBrainz, and eventually on `mic.lt` itself.
- We added Lithuanian folk acts with similar popularities, ages, and other properties currently not featured on `mic.lt`.

This is in the making with Gabija. We want to share with them our findings so that they understand the gravity of the problem and value of our offering.

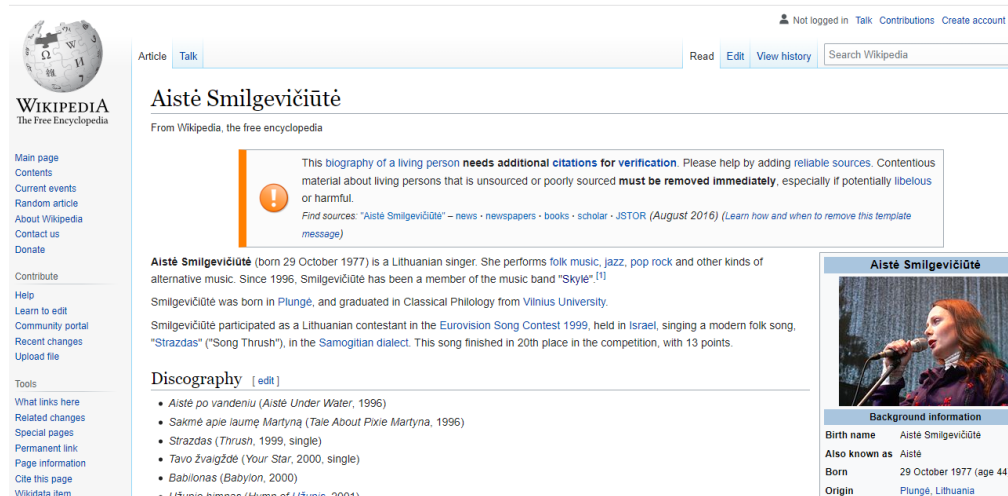


Figure 1: Even though mic.lt has reliable information on Aistė Smilgevičiūtė, it is not communicated automatically to global knowledge hubs.

See *Aistė Smilgevičiūtė* on Wikidata | Wikipedia | MusicBrainz | Europeana | Virtual International Authority File | ISNI. Even though she is a well-known artists, here representation on the information backbone of the internet is very imperfect.

- Less known Lithuanian artists are often completely unknown to these systems, which means that the autonomous systems of Apple, Spotify or Google cannot find reliable information about them.
- Artists who do not have a professional data team may never find out why they are not placed on playlists or not recommended to listeners even in the town where they live, let alone abroad.

Feasibility planning

We assume that the better documented an act is, the most likely that it has many followers and listeners. While this seems to be a trivial hypothesis, we want to measure precisely how many more followers they may have, controlling all other factors equally.

We want to compare artists of similar age, recording history, recency of new releases, sex, or gender. Successful artists with an excellent promotional or label background will have a better digital presence. However, we will measure as much as possible how much of their better presence is a consequence of their earlier success and how much their better digital presence is a cause of their higher relative visibility.

In other words, we know that artists with a Library of Congress entry, a perfect Wikipedia page, and a biography on Spotify are likely to be at an advantage in finding and retaining an audience on Spotify. However, we would also like to measure how much we can help a less successful artist by improving a low-quality Wikipedia page and connecting it with mic.lt, MusicBrainz, and the Library of Congress. This will help us, and it will help other Lithuanian organizations to make a proper cost/value analysis of the necessary information investments into better promoting the Lithuanian music repertoire.

Service offer for the Lithuanian Music Information Center

It is not trivial for music professionals who are not familiar with web technologies why we want to connect a Lithuanian folk singer to the Library of Congress or the Martynas Mažvydas National Library of Lithuania.

In 2022 large global services like YouTube (owned by Google) or Spotify, and Apple Music use automated data collection and artificial intelligence to place music automatically into the right audiovisual stream, playlist or recommend it to the right person. Such recommendations rely on all information available on the web about the artist: partly because the artist or the management provided this information to these services and partly because global knowledge centers already have information about the artists.

For example, there are many artists named “Rain”.

Let’s make an illustration here.

See issue 21

If a new Lithuanian band chooses this name, it is almost sure that at one point, one web service or another will confuse the band with another artist named “Rain”. That will lead to improper recommendations and a negative feedback loop (fans of the Korean artist Rain will be dissatisfied and give negative feedback.) Sometimes they will receive their royalties very late. They may have to do extra paperwork to get a small amount of money – and the paperwork may cost more than the amount outstanding. Or their money may end up on the bank account of another band somewhere else in the world.



Figure 2: Libraries have been familiar with name ambiguity for centuries.

Libraries have been familiar with this problem for centuries. At least 32 literary authors use the name “James Campbell.” Libraries have been building global disambiguation services for decades that ensure each James Campbell is placed on the correct shelf or recommended to the right reader. Such services are also used by collective management agencies like LATGA and AGATA to pay out royalties after a radio play to the correct “Rain” band. However, these services are unavailable for self-releasing artists on YouTube or Spotify because streaming is not collectively managed (AGATA has no legal mandate on these platforms, and LATGA has a very limited mandate.)

We want to help the Lithuanian Music Information Center to connect its database to such linked open data resources (often referred to as the semantic web or the web 3.0) for machine-to-machine synchronization. We want to improve the documentation of Lithuanian artists on the mic.lt website and everywhere else in work.

1. We would like a license to make a copy of the mic.lt public database.

2. We want to connect each artist in the mic.lt database to other reliable knowledge points, such as the Library of Congress, the Deutsche Bibliothek, and Dbpedia (a synchronization tool among 'official libraries' and Wikipedia, among other reliable sources.)
3. We would like to suggest further improvements to the mic.lt database, such as links to the artists' official Spotify or YouTube channel and potential improvements to the current description of the artist's work.
4. We would also like to use the mic.lt as a reliable information source to update the Dbpedia and MusicBrainz entries of the artists automatically. The first visible results will be better Google search results and improved web pages in both the English and the Lithuanian (or German) language versions of Wikipedia.
5. We will try to place as much of this information on the Google Knowledge Graph (for improved Google search results), the knowledge graphs of naming authorities, and library systems as possible. This is important because many services, like Apple Music or YouTube or Spotify, rely on these information backbones of the world wide web.
6. We want to measure the impact week by week: **do artists better placed on web 3.0 see an increase in traffic?** In payments?
7. We want to automatically create a **health check** and provide feedback to every artist in the database for potential problems we encounter while connecting web services and cross-checking their contents.

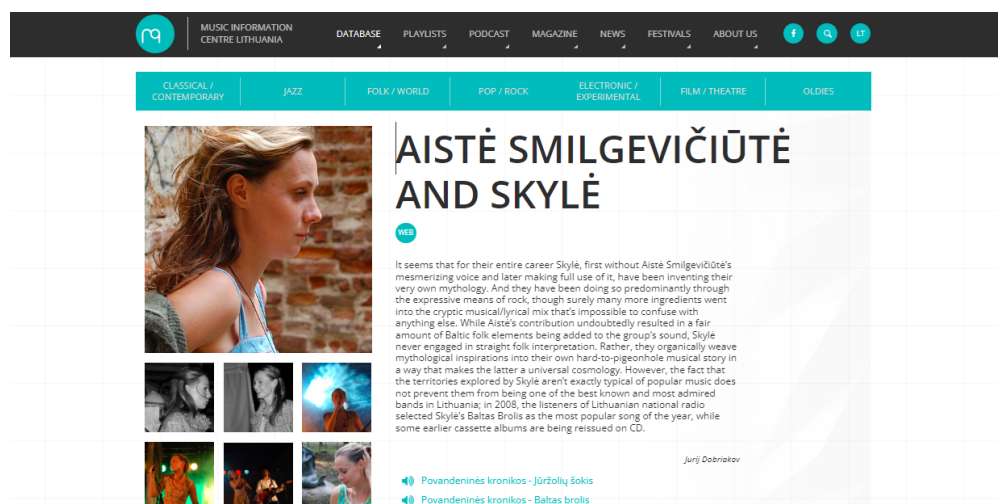
These problems, like name ambiguity, missing metadata, etc., may result in becoming invisible on the service or remaining visible but not getting paid.

Cooperation Requirements

We would like the right to reuse and modify the current public entries of mic.lt database. More precisely, we would like to have a Creative Commons CC0 license for the database.

This will require a revision from Bori

issue 22



- The `mic.lt` website contains information that is authored. For example, *Jurij Dobriakov* wrote an encyclopedical article about Aistė Smilgevičiūtė. We do not want to reuse without modification the work of *Jurij Dobriakov*. However, we want to get the right to reorganize the information in his work, for example, to clarify that *Aistė Smilgevičiūtė* is known to be a woman, and check that this information is correctly placed to Wikipedia or MusicBrainz. We do not think that we need a license for this, however, we may need a license to place this information into a new, derived database.
- We would like to place a quotation from *Jurij Dobriakov*'s article into Wikipedia or MusicBrainz, linking and citing `mic.lt` as the source of the information and, if necessary, *Jurij Dobriakov* as an author. We do not think that we need a license for this, however, we may need a license to place this information into a new, derived database.
- Whenever a reusable photographic image is missing from a global library, we would like to place one of the images of *Aistė Smilgevičiūtė* into this library. Global libraries often contain low-quality, amateur images of the Lithuanian artists, because they do not find legally usable high-quality images. In our view, it is in the best interest of *Aistė Smilgevičiūtė* and *mic.lt* to present Lithuania's music heritage and offering in a better light. However, we understand that including photographic images into a database may be legally more challenging. If the *mic.lt* images were created as work for hire, no license may be needed from the photographer. However, it is also possible that an explicit permission would be required from the photographer(s). In this case, of course, we would only include in our derived database the photographic images (with proper attribution and credit) if *mic.lt* can provide us with both the information and the permission.
- We need to know if *Jurij Dobriakov* must consent to what we are planning to do. We also need to know if *Aistė Smilgevičiūtė* needs to consent to the steps above. We believe that this is not the case, but unknown to us `mic.lt` may have agreed with them otherwise.
- The CC0 license is a pre-requisite to place parts of the information into global libraries, Wikidata and Wikipedia, or MusicBrainz. We want to get the license to synchronize content: give back information that is missing or inaccurate on `mic.lt`, and provide information from `mic.lt` when it is missing from global services.
- We need the right to modify the database. The database in its current form does not follow the W3C (world wide web consortium) standards on web 3.0, i.e., automated data exchange among machines. We would create a web 3.0 version of the existing database. The CC0 license allows the modification of the current database.
- In every case, we will refer to `mic.lt` as the **web 1.0 version** of the database. We want to give proper attribution to `mic.lt`. For example, whenever we place some information from the existing `mic.lt` database onto Wikipedia pages, we will link back to the `mic.lt` original page as the source of information. The CC0 license would allow us to link and “quote” the `mic properly.lt` information on global services.
- We do not want to “own” the resulting, improved, **web 3.0 database**. We would be delighted to set up a joint SPARQL server (which can be automatically connected to global libraries) to contain the improved `mic.lt` and Listen Local Lithuania joined database. We think that the cost of maintaining such a server would be very low, and we would happily assume 100% of its expenses for 2023 and later share these (small) costs with `mic.lt` (or find a suitable Creative Europe grant for continuity.)

Offer to Artists

We see that the musicians have contact details in the database. We would not like to share these details with anybody; however, if necessary, we would like to ask explicit consent from these artists to submit their information from `mic.lt` to the global libraries. It is unclear to us what rights the musicians have on the database.

This is requires a separate review with Bori. This is a privacy / GDPR issue, too.

See issue 23, only after issues 21 and 22 are resolved.

1. Regardless if the artist's explicit consent is needed to use the currently available information on `mic.lt`, we would like to contact them (or their heirs) anyways to inform them about our research and prototype.
2. We are confident that our project would improve their digital visibility even without their involvement and knowledge. However, we could do a lot more with their involvement. Our limited prototype budget can fix only a limited amount of data problems. However, we can send them a list with educative examples of what we see about their web presence and red flag potential problems that may make them invisible or lead to payment problems.
3. We would like to apply (preferably together with `mic.lt`) for a Creative Europe funding to provide non-profit data service to them with a small co-payment. (Usually, some level of co-payment is necessary.) To plan such a granted project, we have to see if they see value in our data health check. We would also like to find out if they or any other organization would be willing to pay a certain amount to fix some of the problems we cannot do in this prototype.