

# CfP

Daniel Antal

2023-10-14

## Table of contents

Summary . . . . .	1
The case of music: there is no commercial and heritage music . . . . .	1
A new data strategy with the mixed use of privately and publicly held data . . . . .	2
Registers & Data Coordination . . . . .	4
Reproducible research and open science . . . . .	5
Expected Impact . . . . .	7

## Summary

We will present a large prototyping project, the creation of the working prototype of the future European Music Observatory and its national pilot, the Slovak Music Dataspace, and the comprehensive Slovak Music Database, as a comprehensive, conceptual framework for connecting privately held and public infrastructure for open science, open policymaking, open data, and privately held cultural heritage and commercial systems. Our aim is to create a large-scale experiment that serves as an empirical hypothesis-testing ground and proof of concept for various ideas that can increase the efficiency of digital curation with better reuse, improved rights management, and a cross-fertilisation between reproducible research techniques and digital cultural heritage management in general. In this large-scale prototyping project, we have an ambitious interdisciplinary goal: to facilitate a better application of parallel work in open (reproducible) science, digital humanities, open data, statistics, and industry.

## The case of music: there is no commercial and heritage music

The digitisation of the enjoyment of culture, then the cultural production, and eventually the way cultural heritage is preserved and made accessible became primarily digital over a very short period, mainly between 1995-2023. The same time brought unprecedented connectivity via the opening of the world wide web in 1991-1993, and the slow uptake of the web of data,

or the semantic web, since 1999 -- where cultural heritage organisations have played the most crucial role.

In the early phases of building our digital infrastructure, people could not imagine and foresee how audiences would enjoy digital culture, form their new habits, or what kind of innovation would form them. For example, when Apple's revolutionary mp3 player came to the market, the late David Bowie prophetically foresaw that access to music would be like access to water. Twenty years later, with digital streaming technologies, Spotify, Apple, Deezer and countless other providers provide subscription access to almost the entire global music repertoire. Any European citizen can have a more extensive catalogue at home than a public broadcaster used to have twenty years ago. The quickly decreasing costs of digitisation and increased connectivity due to exponential bandwidth and semantic connection growth brought unimaginable use cases and earlier unforeseen risks.

In the early decades of digitisation and interconnection, we built institutional and infrastructure solutions that could be more optimal for the demands of 2023. The European Union itself finances many parallel digital infrastructures and solutions. And we see an almost absent coordination between private and public digital infrastructure, which leads to underutilisation of investment and unnecessary cost.

In the early decades of digitisation and interconnection, we built institutional and infrastructure solutions that could be more optimal for the demands of 2023. The European Union itself finances many parallel digital infrastructures and solutions. And we see an almost absent coordination between private and public digital infrastructure, which leads to underutilisation of investment and unnecessary cost. Music and its performances are covered by an extensive intellectual property protection for up to 150 years. Copyrights are recognised 70 years after the death of the longest-living co-author. Neighbouring rights or sound recording copyrights are awarded to the performers who interpret the work and producers who invest in recording it for decades. The administration of these rights and their monetisation created a truly global private infrastructure that carries music cultural objects and describes them extensively about their content properties and the persons and enterprises involved in their making. This global digital infrastructure is not connected to the European digital heritage infrastructure. While about a hundred million recordings of tens of millions of music works are available on private but privately freely accessible infrastructure, Europeana only refers to fewer music works than uploaded daily to Spotify. The concept paper of ECCH does not refer to the word "music".

## **A new data strategy with the mixed use of privately and publicly held data**

"A register aims to be a complete list of the objects in a specific group of objects or population." (Anders and Britt 2007) We are planning music industry registers where the objects are *music works* and *sound recordings* (in statistical terms, music products), and the populations are *music authors*, *music performers*, *groupings of performers* (as the majority of the musicians

perform, record, release in groups, ensembles, orchestras), *record labels* (which may be formal and informal businesses) and *music publishers* (enterprises.)

We chose a statistical framework for this task, because the data and metadata infrastructure of statistics has been globally standardised for decades. We believe that if we want to create comprehensive statistics on the users (audience) of music, the performers and creators or music, or the available or used music works, live and recorded performance, we need to create comprehensive registers. Reference to these registers in scientific inquiry or digital curation would greatly increase transparency.

Although Europeana Sounds, based at the British Library, continues to be an aggregator for audio and audio-related material for Europeana, a quick search for recorded music in Europeana [Music Art/activity of creating art using sound](#), brings up only 14,252 hits.

More music becomes available digitally on a day with rich, machine-readable metadata than the entire body of music that can be found in Europeana, which in itself aims to be an umbrella of European heritage organisations. For comparison, CISAC's global database, which is a shared creation of 225 creator societies in 116 countries in the private sector, contains information with pristine metadata about 52 million music works identified with the ISO 15707:2001 standard International Standard Work Code. These societies hold a sound recording, and in many cases, machine-readable or only human readable notation of the music work in copy. Less than 0.01% of them is even referenced in Europeana, and that aims to be an interoperable aggregator umbrella of European cultural heritage. To make another comparison: in 2022, 213,000 artist, new creators, performers of authentic folk or classical music, and many others made *at least ten* new recordings available—with dozens of features that enable AI-driven algorithmic recommenders to seek audience for them globally from the day of release—, resulting on a growth of more than 2 million available music recordings. These recording can be found, enjoyed, listened to legally and even free, if the user is willing to listen to commercials, too. Creating joint music registers could increase the visibility of European music by at least 1000x times. We acknowledge that there are certain folklore or liturgical traditions of music that do not encourage or even forbid the recording of a musical performance, but such music works and their performances are exceptions to the ubiquitous presence of digital music objects and their various information or metadata objects.

The European Data Strategy, a high-level strategic policy that already resulted in several pan-European pieces of new data governance legislation, wants to foster the re-use of public information (this is the aim of the Open Data Directive) and public-private exploitation of data, with even increasing the use of privately-held data for public purposes.

Our current pilot project aims the creation and harmonisation of Slovakia's population and business registers with the registers of the music industry and music libraries and heritage organisations with the creation of a Slovak music data space and a semantically interlinked Slovak Comprehensive Music Database. Our goal is to connect information about every natural and legal person who created music as an author, or as settled performer of music in the current

territory of the Slovak Republic, which was part of Czechoslovakia between 1918-1992, and earlier part of the Hungarian Kingdom and Austria-Hungary.

Our first step is the creation of the Slovak Music Industry Register, which would connect legal persons, including non-profit heritage organisations, collective rights management organisations, and for profit music publisher and record labels, following the *United Nations Guidelines on Statistical Business Registers. Final draft prior to official editing* ('United Nations Guidelines on Statistical Business Registers. Final Draft Prior to Official Editing' 2020) which is heavily based on the former UNECE *Guidelines on statistical business registers* (UNECE 2015) and the EU-relevant *European business statistics methodological manual for statistical business registers. 2021 edition* (Eurostat 2021). We intend to connect this collaborative, public-private register with a structural business register satellite according to these guidelines to create an unambiguous data coordination tool among already harmonised state resources—such as the state registry of companies that includes commercial and non-commercial music organisations, or the population of living artists.

To achieve this ambitious goal, our research and innovation project (supported by the European Commission) signed a memorandum of understanding with the Cultural Ministry of the Slovak Republic, the Institute of Cultural Policy, and initiated a collaboration with all relevant music institutions, libraries, businesses in the country, and the national statistical authority (Ministerstvo kultúry SR and Open Music Europe 2023).

## Registers & Data Coordination

Our approach is rooted in statistics. In the era of big data, data and statistics are often confused; statistics, however, is the best scientific tool for humans to see the bigger picture and patterns in too much data.

A *statistical register* is a continuously or regularly updated set of objects for a given population. It contains information on the identification and accessibility of population units as well as other attributes which support the surveying process of the population. It serves as a constantly updated list of potential data sources: people (for example, performers or authors of music) or enterprises (for example, music labels, concert halls, music publishers). They can be surveyed in a census, when all people or enterprises must provide data, or in a sample survey, when some of them are randomly selected. The statistical register is both a coordination tool for data collection (everybody is found who should provide data), and an important data quality management tool (we know if somebody was not found, how it will distort our resulting datasets.) For example, **Statistics Norway** applies about 100 statistical registers. Our primary concern in Slovakia is the creation of a music business register, because this could provide indicators for the public policy-level and institutional/enterprise level implementation of the Slovak cultural strategy, with wide-ranging impact on the transparency of digital curation.

We want to create reliable statistical data on the growth of the body of music work and sound recordings made in Slovakia in a way that is transferable to any country that uses the UN guidelines on statistical production. We want to produce statistics on the sociological composition of performers and authors of music, such as the total number, gender ratio, age, or the institutions that support the creation, recording or preservation of music, such as the average cost of creation, or the carbon footprint of releasing, keeping in circulation or archiving music. We believe that access to such information is an important starting point of creating a good curation strategies or data stewardship for the public or business-oriented (market) research of musicology, music economy, cultural policy or music sociology.

The novelty of our approach lies in a data coordination between public (state) organisations which had been already created in the last three decades, and the private sector, which exists only in experimental form for about ten years.

Slovakia has already made a significant and exemplary investment in creating a satellite account system for the cultural and creative industries (Horecka and Némethová 2022). It already produces high-quality statistical information about architecture and advertising, yet it remains particularly challenging to extend this approach to other cultural sectors. Our aim is to separate the existing audiovisual aggregate statistics to usable, actionable music, film and television statistics that can form the basis of good curation and data stewardship policies in these important domains. The necessary steps towards this lie in the creation of an ex ante harmonised, public and private survey program extending the current xxxxx with a sufficiently large and well-coordinate sample survey of music (social and for-profit) enterprises, a well-harmonised cultural access and participation population sample survey, and potentially the addition of a private labor force survey that tackles the high level of informality among music sector professionals.

Surveying with questionnaires is costly for both respondents and the institutions that administer the surveys, therefore, whenever possible, we would like to coordinate the existing privately-held data sources. For example, we do not need to survey music publishers and authors of music about new works, if we can access for statistical purposes the register of Slovak music works in SOZA: this is a far more timely and accurate source of the information.

## **Reproducible research and open science**

Our approach is rooted in statistics, but it is not purely statistical. Designing an excellent curatorial strategy or research plan almost always involves some statistics, even if the research is qualitative. It is easier to imagine musicological research into Slovak music by being familiar with the scope, breadth and size of music that can be connected to the Slovak language, the territory and regions of Slovakia or the people who have lived or worked in the country. We do not only want to provide access to statistical information; we see the creation of comprehensive

registers and their summary statistics as a necessary first step into open scientific research or digital curation.

A harmonisation of public and private registers is a great step towards making data, information or knowledge findable, technically accessible and interoperable, and to make it machine-usable, notwithstanding legal, ethical or legitimate business concerns to limit access and use. Ideally, with a good and ethical research plan and appropriate data and copyright protection, researchers should be enabled to find the typical musical harmonies in a well-framed body of work concerning time, region, or similarity to some other body of music via machine-readable notation of the works; recognise patterns in the use English and Slovak works in the connecting lyrics; test hypothesis about the changing female participation ratio in live and recorded performance that meet certain success criteria or to benchmark their music festival’s carbon footprint to other temporary festival infrastructure’s footprint. The set of such ideal and desirable goals is infinite. Resources and limitations in legal barriers of conflicting interests are endless, but the coordination of privately- and publicly-held information with strict metadata standards, shared namespaces, and using the existing state infrastructure is a necessary first step to consider the costs and benefits of realizing any such research goal.

Our work is connecting four different guidelines, which are not standards or protocols, but which are well supported by them: GSIM from statistical innovation, FAIR from open science, OPA from open policy analysis, and the tidy data principle, which is inspired the renewal of the open source R statistical environment via tidyverse. While we offer to develop a solution in the R language, the research data management principles that we develop can be easily replicated in more general languages used for data science, for example, in Python or C++. Our aim is to support individual researchers or small groups to work with our data space without reliance to extensive data or knowledge engineering resources to support them in their workplaces.

GSIM is a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information that are used in the production of official statistics (information objects). It creates a shared vocabulary between the SDMX and DDI. SDMX stands for the Statistical Data and Metadata eXchange: it is an international initiative that aims at standardising and modernising (“industrialising”) the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries. SDMX has an ISO standard that is well harmonised with W3C linked open data vocabularies. The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences. Last, but not least, the tidy data principle connects some fundamental principles of computer science with statistics. It is widely used in the recent development of the R language itself, but as a principle, it is applicable for any computer language used for a programmatic work with tabular data, including most business ledgers, accounts, books, or institutional registers.

We hope to achieve a high compliance with SDMX with the creation of “experimental statistics” together with the Statistical Office of the Slovak Republic (SOSR). SDMX-

compliance would give the users of our data space the highest level of machine-readable interoperability and quality for statistical, bird-eye view oversight. Implementing at least aspects of DDI would allow the reproducible, automated production of such statistics, or offer machine-actionable access to the information objects for research. Both SDMX and DDI covers a very wide range of automated statistical and research processes and potential information carriers; we will subjectively select elements that are practically needed for a Slovak music data space, and obviously disregard for example processes relevant for health sciences.

FAIR and OPA are not standards or protocols either, but guidelines. FAIR is widely accepted in open science, and it is mandatory to apply some elements of this guidelines in every EU-funded research. OPA is a newer known standard that aims to apply transparency and reproducibility guidelines of social sciences (covered by FAIR) for public policy analysis, thus making policy analysis evidence-driven and interoperable with open science.

## Expected Impact

On the level of **public policy**, we aim to provide indicators and benchmarks for the sectorial Stratégia kultúry a kreatívneho priemyslu Slovenskej republiky 2030 (Ministerstvo kultúry Slovenskej republiky 2023)—henceforth: Slovak CCI Strategy, the Slovakia’s Vision and Development Strategy 2030 - a long-term strategy for sustainable development (Ministerstva investícií, regionálneho rozvoja a informatizácie Slovenskej republiky 2020)—henceforth: Slovak Strategic SDGs, which has significant cultural elements, and the 2030 Digital Transformation Strategy for Slovakia (Ministerstva investícií, regionálneho rozvoja a informatizácie SR 2019a, 2019b)—henceforth: Slovak Digitisation Strategy. Throughout the project, we will follow the Eurostat guidelines on creating new indicators (Eurostat 2014, 2017; Kotzeva et al. 2017) and we aim to substantially increase the knowledge base for cultural policies that were inventorised in cooperation with the European Commission’s Structural Reform Support Service (SRSS) in 2020 (KEA 2020). Our Memorandum of Understanding, as far as we know, represents the first such national effort to commit national, state institutions and open science partners to the principles of Open Policy Analysis.

We aim to make our work replicable at least throughout the European Union and the European Economic Area, therefore we constitute our data space under the umbrella of the European triple transition strategic policy agenda, which can be connected to the aforementioned Slovak public policies and mapped to any other country that applies this framework. After the COVID-19 pandemic, policymakers in Europe increasingly discuss the strategic public policy agenda under the triple transition term, which refers to approaches in climate issues (“green transition”), technology (“digital transition”) and the underlying shape of its society (“social transition”). The “triple transition” is highly relevant to our work because these horizontal policies guide public investment, have new rules on private investment spending, and foresee changes in economic and tax policies. The EU’s cultural programme Creative Europe already follows the triple transition framework. (**muench\_et\_al\_2022?**).

On the level of business policy, we follow an inventory of problems identified in 2020 in the xxxxxx. We have created an ambitious program to xxxxxx,

Last, but not least,

Anders, Wallgren, and Wallgren Britt. 2007. *Register-Based Statistics. Administrative Data for Statistical Purposes*. 1st ed. Chichester:United Kingdom: John Wiley & Sons Ltd.

Eurostat. 2014. *Towards a Harmonised Methodology for Statistical Indicators — Part 1: Indicator Typologies and Terminologies*. 2014th ed. Vol. 1. Towards a Harmonised Methodology for Statistical Indicators 1. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-14-011>.

———. 2017. *Towards a Harmonised Methodology for Statistical Indicators — Part 3: Relevance for Policy Making*. 2017th ed. Vol. 3. Towards a Harmonised Methodology for Statistical Indicators 1. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-17-007>.

Eurostat. 2021. *European Business Statistics Methodological Manual for Statistical Business Registers. 2021 Edition*. Manuals and Guidelines. Luxembourg: Publications Office of the European Union. <https://doi.org/10.2785/093371>.

Horecka, Horecká, and Renáta Némethová. 2022. ‘The Summary Results of the Satellite Account on Culture and Creative Industry Ofthe Slovak Republic (2015-2020)’. Edited by Ján Haluška. INFOSTAT – Institute of Informatics; Statistics. <https://www.culture.gov.sk/wp-content/uploads/2019/12/The-summary-results-of-the-Satellite-Account-on-Culture-and-Creative-Industry-of-the-Slovak-Republic-2015-2020.pdf>.

KEA. 2020. ‘Cultural Industries in Slovakia. Report on the AS-IS Situation. Analysis and Recommendations to Boost Growth in Cultural Industries in Slovakia’. Brussels: KEA European Affairs. [https://www.culture.gov.sk/wp-content/uploads/2020/11/AS-IS\\_sprava.pdf](https://www.culture.gov.sk/wp-content/uploads/2020/11/AS-IS_sprava.pdf).

Kotzeva, Mariana, Anton Steurer, Nicola Massarelli, and Mariana Popova, eds. 2017. *Towards a Harmonised Methodology for Statistical Indicators — Part 2: Communicating Through Indicators*. 2017th ed. Vol. 2. Towards a Harmonised Methodology for Statistical Indicators 1. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-17-001>.

Ministerstva investícií, regionálneho rozvoja a informatizácie Slovenskej republiky. 2020. ‘Vízia a stratégia rozvoja Slovenska do roku 2030 - dlhodobá stratégia udržateľného rozvoja [Slovakia’s Vision and Development Strategy 2030 - a Long-term Strategy for Sustainable Development]’. <https://mirri.gov.sk/wp-content/uploads/2021/01/Slovensko-2030.pdf>.

Ministerstva investícií, regionálneho rozvoja a informatizácie SR. 2019a. ‘2030 Digital Transformation Strategy for Slovakia’. Ministerstva investícií, regionálneho rozvoja a informatizácie SR. <https://mirri.gov.sk/wp-content/uploads/2019/10/SDT-English-Version-FINAL.pdf>.

———. 2019b. ‘Stratégia digitálnej transformácie Slovenska 2030. [20230 Digital Transformation Strategy for Slovakia]’. Ministerstva investícií, regionálneho rozvoja



- a informatizácie SR. <https://mirri.gov.sk/wp-content/uploads/2019/06/Strategia-digitalnej-transformacie-Slovenska-2030.pdf>.
- Ministerstvo kultúry Slovenskej republiky. 2023. ‘Stratégia kultúry a kreatívneho priemyslu Slovenskej republiky 2030. [Strategy of the cultural and creative industries of the Slovak Republic 2030.]’. [https://strategiakultury.sk/wp-content/uploads/2023/06/Strategia\\_KKP\\_SR\\_2030\\_po\\_shvaleni\\_.pdf](https://strategiakultury.sk/wp-content/uploads/2023/06/Strategia_KKP_SR_2030_po_shvaleni_.pdf).
- Ministerstvo kultúry SR, and Open Music Europe. 2023. ‘Memorandum o porozumení o využití výsledkov analýz otvorených politík v kontexte slovenského kultúrneho a kreatívneho priemyslu a sektorových verejných politík v spolupráci s konzorciom pre výskum a inovácie s názvom OpenMuse. [Memorandum of Understanding on utilizing the Open Policy Analysis results of the OpenMuse Research and Innovation Consortium in the context of Slovak cultural and creative industries and sectors’ public policies]’. <https://www.crz.gov.sk/zmluva/7645338/>.
- UNECE. 2015. ‘Guidelines on Statistical Business Registers’. United Nations Economic Commission for Europe. [http://www.unece.org:8080/fileadmin/DAM/stats/publications/2015/ECE\\_CES\\_39\\_WEB.pdf](http://www.unece.org:8080/fileadmin/DAM/stats/publications/2015/ECE_CES_39_WEB.pdf).
- ‘United Nations Guidelines on Statistical Business Registers. Final Draft Prior to Official Editing’. 2020. United Nations. [https://unstats.un.org/unsd/business-stat/SBR/Documents/UN\\_Guidelines\\_on\\_SBR.pdf](https://unstats.un.org/unsd/business-stat/SBR/Documents/UN_Guidelines_on_SBR.pdf).