Indicators in Open Music Europe

Daniel Antal

2023-09-18

Table of contents

Defining the Data-to-policy Pipeline

`Open Music Europe` offers a 'data-to-policy' pipeline, which extends the music data pipeline to evidence-based business and policy administration. A data pipeline is a method in which raw data is ingested from various data sources and then ported to data store, for further analysis, in this case, to an open, shared, collaborative music observatory. We extend this pipeline using reproducible research techniques, a novel application of the *Open Policy Analysis Guidelines*, and good statistical practices to support evidence-based policy analysis, scientific music research and sound business strategy building. In this last leg of the pipeline, we emphasise usability for our project's target audiences and good documentation practices. We want to ensure that our data is high quality and well understood to support robust and correct business, scientific or policy conclusions. This usually makes processing the data into an indicator, or a set of indicators, which are often displayed as a scoreboard, dashboard,

or as a part of a standardised business or policy report.

Data coordination and ingestion

A data pipeline is a method in which raw data is ingested from various data sources and then ported to data store, for further analysis, in this case, to an open, shared, collaborative music observatory.

We want to create a data production method that makes already available public data more usable for the purposes of music businesses, researchers, and policymakers and links further data to fill data (information) gaps. The first leg of the data-to-policy pipeline requires the creation of a statistical infrastructure, a data observatory, that can receive processed and unprocessed public and private data, link them, and offer a data pipeline for newly collected data, too.

From data to indicators

The use of indicators is widespread in the economics, business, social aspects, or public policies of music and the broader cultural and creative industries agenda. A statistical indicator summarises a key issue or phenomenon derived from a series of observed facts; it is a generalisation of the information in the data. It is often compared metaphorically to traffic signs that help travelers to arrive to their destinations safely. The knowledge we gain from indicators is used to uncover social, environmental or economic phenomena and establish connections between them. It also provides a basis for influencing and controlling such phenomena on a business level (for example, key performance indicators) or the level of public policy (for example, *ex ante* and *ex post* evaluation indicators.)

> A statistical indicator is a summary measure related to a key issue or phenomenon and derived from a series of observed facts. [Kotzeva et al. (2017), pp. 7][1]

A `statistical indicator`, or simply, `indicator`, is the representation of statistical data for a specified time, place or any other relevant characteristic, corrected for at least one dimension (usually size) so as to allow for meaningful comparisons. It is a summary measure related to a key issue or phenomenon and derived from a series of observed facts.

Statistical indicators, or, in short, indicators, are derived from accounting systems and statistical data. Accounting systems are coherent and integrated accounts, balance sheets and tables based on a set of agreed rules. Statistical data is also collected based on agreed and coherent rules.

---

[1]The definition of the statistical indicator originates from the Regulation 99/2013 of 15 January 2013 on the European statistical programme 2013-17 (Annex I,1: Indicators), which went through several modification and it is no longer in force (European Parliament and the Council 2013; EUR-Lex 2018). In a slightly more verbose format it is part of Eurostat's Glossary.
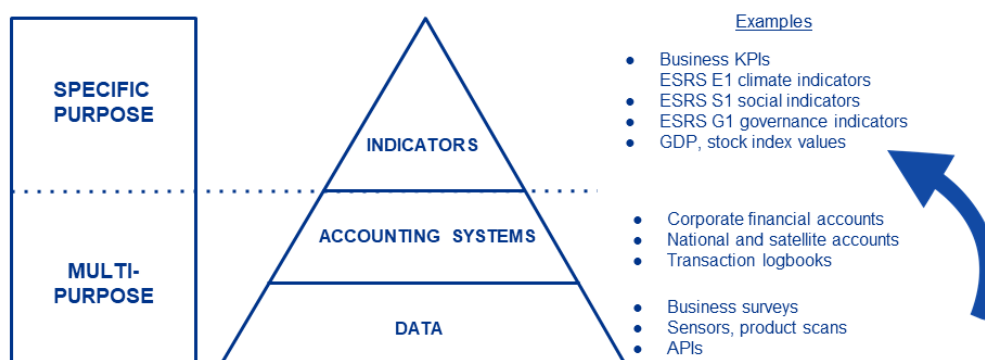
SPECIFIC PURPOSE

MULTI-PURPOSE

INDICATORS

ACCOUNTING SYSTEMS

DATA

Examples

- Business KPIs
  ESRS E1 climate indicators
- ESRS S1 social indicators
- ESRS G1 governance indicators
- GDP, stock index values

- Corporate financial accounts
- National and satellite accounts
- Transaction logbooks

- Business surveys
- Sensors, product scans
- APIs

Figure 1: The statistical infrastructure of indicators, DOI: https://doi.org/10.6084/m9.figshare.23600571

The aim of statistical data collection is the creation of sets of data that can be generalised well; for example, when we collect data about people, it includes data on people from all relevant age or gender groups. `Open Music Europe` is using and developing data collection from primary sources, such as surveys and ephemeral big data, or taking samples of large, pre-existing but biased datasets, such as royalty accounts.

From indicators to policies

Indicators should be used as traffic signs, and not as route plans. Indicators do not give an explanation to a problem or reveal a solution; they must be interpreted in the context of domain-specific knowledge. While the word *policy* is very often used in the context of public policies, we use it in a more general meaning: a course or principle of action adopted or proposed by an organization or individual. `Open Music Europe` does not offer tailored data analysis for organisations or individuals, it offers indicators and other data products that can help the setting of policies for public policy, for-profit and social music enterprises, and hopefully even for individual creators. We would like to do this with placing indicators into reproducible reports. These reports contextualize our indicators, and end the data-to-policy pipeline.

> Unlike the data from which they are derived, indicators are used for specific purposes. The specific purpose of an indicator is determined by its context, the questions it seeks to answer, and the party or parties asking these questions. (Eurostat 2014, 1:p7)

Indicators are used for specific purposes to support decisions with generalised information derived from the data. The specific purpose of an indicator is determined by its context, the questions it seeks to answer, and the party or parties asking these questions.

There are many conceptualisations on how policies should be set for public organizations or businesses; a review of policy cycles is beyond the scope of this paper. For presentation purposes, we use the policy cylce framework of …. adopted by Eurostat for indicator design guidance.

In an evidence-based policy cycle indicators should be provide different factual information. Different indicators may be needed in the first, `Concept` phase when a business or public body is seeking alternatives to solve a problem: for example, to increase royalty revenue for their artists or assets, and in the final `Outcome` or `Impact` assessment. While impact assessment is often associated with public policy, the current sustainability management techniques require music businesses to implement processes that reduce, for example, fossil fuel use, and eventually have a positive impact (in quantitative terms, a negative change) on greenhouse gas emissions.
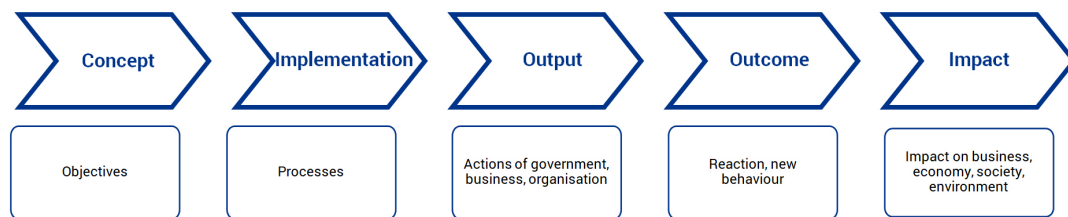

Figure 2. Policy evaluation cycle with indicators, DOI: 10.6084/m9.figshare.23498441

Following the data pillars of the *Feasibility study for the establishment of a European Music Observatory : final report* (Commission et al. 2020), we foresee the indicators made in our sample reports with the following topics:

- **Report on the European Music Economy**: Support stakeholders with economic questions, such as taxation, price-setting and economic exploitation, and advocating for policies that better support the development of the industry or its parts like publishing (Open Music Europe 2023e).

- **Report on Music Diversity and Circulation in Europe**: Local content regulations, maintaining the diversity of music circulation in Europe, and other aspects of diversity (Open Music Europe 2023b).

- **Report on Music, Society, and Citizenship in Europe**: Inform about the societal aspects of music, amateur practices, education, and sustainability (Open Music Europe 2023d).

- **Report on Music Innovation & Technology in Europe**: Novel data-driven applications (Open Music Europe 2023c).

Interactive communication of indicators and Open Policy Analysis

Indicators are most important in communicating data in a business or public policy context. To discuss this communication, it is helpful to distinguish between simple one-directional information transmission and communication as an interactive process. The informational

or one-directional communication of indicators focuses on information transmission between a sender and a receiver. The interactive communication of indicators means that the message is actively constructed by both the sender and the receiver(s), who exchange information in a feedback loop.

> Like the mediators who strive to communicate between different parties to find their point of agreement, communication through indicators allows to overpass boundaries between policy, science, statistics and public debate. Thus indicators are able to reveal various phenomena to different user groups — policy makers, researchers, citizens, journalists, statisticians, specialists in different domains. Reaching various types of audiences through their communication function, indicators can ensure that statistics are usable and useful (Kotzeva et al. 2017, 2:p7.).

Going back to the often used traffic management example, a simple one-directional indicator is a speed limit sign on a static road sign; a more interactive version uses sensors and cameras to measure the actual traffic density and to recommend a speed at which vechicles will not jam the road.

Interactivity allows for exchange of ideas and a shared creation of the intended content of the message that the planned indicator conveys. This requires user involvement in the construction of indicators and indicator sets. The developed indicators need to be based on a wide consensus reached through a dialogue involving as many potential user groups as appropriate. This way they can have their say on the relevance of the statistical message. For example, on Digital Streaming Platforms (DSPs) a music creator or label is facing with hundreds of hidden prices. One of our aims is to create price indicators that help a music label, distributor or a band to direct marketing efforts towards more valuable streaming segments. We need to create a price indicator that is as simple as a stock exchange index which is often followed by small and institutional investors alike.

> PRINCIPLE 11 Relevance: European Statistics meet the needs of users. 11.1. Procedures are in place to consult users, to monitor the relevance and value of existing statistics in meeting their needs, and to consider and anticipate their emerging needs and priorities. Innovation is pursued to continuously improve statistical output. 11.2 Priority needs are being met and reflected in the work programme. 11.3 User satisfaction is monitored on a regular basis and is systematically followed up.

The *European Statistics Code of Practice* (European Statistical System Committee 2018) see the *Quality Assurance Framework of the European Statistical System* (European Statistical System (ESS) 2019, 45–46) goes a step further in interactivity. The idea of interactivity in the statistical indicator design process originates in the the European Commission's *more-than-information approach*: „Communication is more than information: it establishes a relationship and initiates a dialogue with European citizens, it listens carefully and it connects to people. It is not a neutral exercise devoid of value, it is an essential part of the political process." The Commission refers here instead of the 'target audience' to 'citizens' because communication is not just a technical exercise of message delivery but part of a

democratic involvement of citizens (Kotzeva et al. 2017, 2:p7.). Similarly, a good business indicator is not only printed in the annual financial or sustainability reports of a company: it is often cited in management discussions or communicated by line managers to workers to direct their focus. To make indicators conversation-starters or arguments in professional discussions, we have to ensure their usability: we not only need to involve potential users in the design phase, but we have to get regular feedback on the usability and professional improvement of the indicators.

Data coordination and ingestion

Curated statistical data

In the last two decades, the availability of statistical data increased to the extent that non-professional users often do not find the information they need. Furthermore, statistical authorities increasingly release data that is less processed and requires statistical and domain-specific know-how to transform into a form that music professionals and researchers, or policy-makers can readily use.

Our seemingly trivial first data pipelines are building access to curated and if needed, re-processed statistical data towards high-quality statistical data sources. The rOpenGov collaboration maintained by the University of Turku, where REPREX is an active developer, provides reproducible tools to access such sources. These tools will be further developed into a statistical ecosystem with a more friendly user interface in WP4.

A good example of these tools will be used and further developed in WP3. The iotables R package on rOpenGov, developed by the leadership of REPREX, provides access to the Eurostat data warehouse's symmetric input-output data. Symmetric input-output and auxiliary tables connect thousands of poorly usable statistical indicators into analytical tables, which can be used for various economic analyses and social or environmental impact assessments. The iotables package helps the user to bring together the necessary data from Eurostat's data warehouse (that includes that from the European Environmental Agency) and perform data processing and algebraic transformations that result in readily-usable indicators and multipliers. This service can be seen as a middleware between data processing and data analysis: most analysts who know how to work with the Leontieff- or Ghosh system to work with this data would spend days just to put together the necessary indicators from the original source.

Reusable microdata and public sector information

The public sector already holds an extraordinary amount of data that can contribute to improving the internal market and to the development of new applications for consumers and legal entities. The Directive 2003/98/EC established a set of minimum rules governing the re-use and the practical arrangements for facilitating re-use of existing documents held by public sector bodies of the Member States; the rules of such data access were updated in

2019 with the adoption of the *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)* (European Parliament and the Council 2019a).

According to the preamble of the Directive, information collected, produced, reproduced, and disseminated within the exercise of a public task or a service of general interest is an important primary material for digital content products and services, with a special emphasis on publicly funded scientific research, such as Open Music Europe, or other Horizon Europe framework funded projects. They may form data monopolies and prevent innovation, start-ups, and SMEs from entering into the Single market; for natural persons and civil society organisations, they may form an insurmountable barrier to control the political process, public policy, or engage in advocating for better policies[2].

Because the amount of data carriers, such as digital documents containing datasets, text, and data visualisations, exceeds what humans can possibly read, curate, and analyse, the directive also emphasises the need to provide access to public sector information in open and machine-readable formats and via APIs[3].

The European open data regime not only sets the normative standards for our projects and its data management, but it also provides an opportunity to tap into a wide array of data sources in public institutions. Building a framework to receive and process large amounts of public information is the task of our WP5 (Open Music Observatory), but it requires much work in the other work packages, too, because domain-specific knowledge and curation are needed to identify the necessary data to be ingested, the procedures to process the data, and to check its quality.

By applying the same framework (principles, legal norms, organisational and technical standards) as the national statistical offices themselves as regards public and private administrative data sources, Open Music Europe can facilitate the data supply for the music sector in multiple ways: - We can engage with statistical offices and government and offer them our innovations to be scaled up using the publicly funded and legally more powerful national statistical frameworks; - Using industry or research funds, we can fill

---

[2]"The preamble of the Directive (EU) 2019/1024 states these aims with justifications (see the legal text for full details):(36) Charges for the re-use of documents constitute an important market entry barrier for start-ups and SMEs. Documents should therefore be made available for re-use free of charge and, where charges are necessary, they should in principle be limited to the marginal costs [...] (44) The re-use of documents should not be subject to conditions. However, in some cases justified by a public interest objective, a licence may be issued imposing conditions [...] (69) For the purpose of ensuring their maximum impact and to facilitate re-use, the high-value datasets should be made available for re-use with minimal legal restrictions and free of charge."

[3]"The preamble of the Directive (EU) 2019/1024 states these further aims(34) To facilitate re-use, public sector bodies should, where possible and appropriate, make documents, including those published on websites, available through an open and machine-readable format and together with their metadata, at the best level of precision and granularity, in a format that ensures interoperability[...] (69) For the purpose of ensuring their maximum impact and to facilitate re-use, the high-value datasets should be made available for re-use with minimal legal restrictions and free of charge. They should also be published via APIs.[...](27)[...] Beside open access, commendable efforts are being made to ensure that data management planning becomes a standard scientific practice and to support the dissemination of research data that are findable, accessible, interoperable and re-usable (the FAIR principle)."

data gaps with complementary data that integrates well with novel data products resulting from the increasing reliance of statistical organisations on third-party data.

The framework for this work is developed in WP5 as the music observatory. The domain-specific elements of this framework must be developed by four work packages.

Data linking and statistical registers

A statistical register typically plays the role of a data coordination tool, integrating data from several sources, both statistical and administrative. This may be done by linking records by means of common identifiers (microdata level), or by using other matching techniques (statistics and indicator level.)

A statistical register is a continuously or regularly updated set of objects for a given population. It contains information on identification and accessibility of population units as well as other attributes which support the surveying process of the population.

Music Industry Business Register

In `WP1` we are developing new indicators for filling the data gaps in the European music economy. For this purpose, we need a framework that efficiently divides data relevant to the music economy and which falls outside of the scope of the music economy. In other words, we need a mapping tool that allows the collection of music industry data separately from non-music industry data.

Our emphasis in `WP1` is the provision of data for valuing zero-price music and other forms of undervalued European content and to provide information on designing better corporate and public policies on music streaming and broadcasting. While Open Music Europe mainly focuses on the publishing and recording side of the music business, we need to build a framework that can support the bigger live performance part of the music industry.

The pilot country in this regard is going to be Slovakia, where at the beginning of the project, we secured to come to a Memorandum of Understanding with the national government. We aim to create a data framework in the music industries that is extensible to the broader cultural and creative industries context and successfully links statistical, government and private data for a better design, implementation and monitoring of Slovak public policies and the business policies of Slovak music stakeholders, including our Consortium member SOZA.

The SKMIBR is an experimental concept of continuously or regularly updating set of objects for a given population: i.e., music enterprises, including freelance music professionals.

Comprehensive Music Registers

In WP2 we are developing new indicators for filling the data gaps in the European music diversity and circulation. For this purpose, we need a framework that efficiently divides data relevant to the local, national or European music stakeholders about music that they consider to be relevant to them and music that is outside of the scope of their interest. For example, suppose they want to monitor the radio quotas defined by the Slovak national legislation for Slovak content. In that case, they have to rely on a list of sound recordings available for broadcasting that are "Slovak" and "Not-Slovak". Or, if they want to monitor the French quotes on new music in a broadcast stream, we must have a reference register of recordings with the creation or release date.

In WP2 we rely on the CEEMID background of the Slovak Demo Music Database, which we will develop into a Slovak Comprehensive Music Database (SKMDb) and replicate in Bulgaria. Such a database fulfils many roles apart from creating actionable indicators of music diversity and circulation on public policy and business levels.

The SKMDb will be a continuously or regularly updated set of objects for a given population: music that is considered "Slovak", or relevant for the Slovak businesses and public policies. Having such a reference list of music works and recordings is an essential prerequisite to understanding, for example, the market share of Slovak music in various foreign markets or the domestic market. Without such a comprehensive registry, local content regulations (which are present in most European markets) cannot be efficiently monitored or enforced; music export strategies cannot be designed, implemented or monitored; and, of course, music labels and distributors are often spending their marketing budgets inefficiently.

Flexible linking of administrative records to sustainability measures

In WP3 we aim to develop music data products to support the triple transition of the music industry to a more digitally apt, environmental and socially sustainable development path. For the first time, after adopting the CSRD Directive, Europe is standardising indicators that can monitor progress towards these goals (and towards more ethical business conducts, too.)

In this case, we need to build very flexible registers that connect the administrative records of music organisations with external sustainability data, for example, connecting their existing management information system with greenhouse gas emission data or data on gender equality measures.

Because WP3 has no methodological budget for developing such a framework, it will rely on partly the symmetric input-output tables and auxiliary sustainabilty tables of the national account systems, and partly on the REPREX background, the *Eviota System*. The Eviota system is developing a similar tool to the sustainability auxiliary (satellite) accounts of the national account system for IFRS-based, European corporate accounting systems used by both non-profit, social and for-profit enterprises.

Generalised frameworks for data linking

In WP4 we want to foster innovative data uses in the music sector. Our aim here is to provide a more general framework for data access and data improvement for novel, innovative data-driven technologies and workflows that can support the music industry.

Data Infrastructure: The Music Industry Business Register Concept

Statistical business registers (BR) are repositories of legal and statistical units to be used for producing business and macroeconomic statistics. They include information on the active population of:

- Enterprises carrying out economic activities which contribute to the gross domestic product (GDP);
- Legal units of which those enterprises consist of;
- Local units;
- Kind of activity units;
- Enterprise groups, incl. all-resident and multinational enterprise groups.

In the current statistical infrastructure, each EU member state has set up a national statistical business register (NSBR) within their national statistical office on the basis of the European Regulation 2019/2152 (European Parliament and the Council 2019b), which updated earlier regulations in this field going back to 1993. These NSBRs connect to the government's tax authorities, company registration institutions like chambers of commerce or government company houses, and courts. They have an essential role in creating business statistics.

NSBRs are not of identical quality across the EU; the motivation behind Regulation 2019/2152 was correcting some differences among member states. Furthermore, a long-known problem of the NSBRs is that they do not cover well very small enterprises (Eurostat 2021, 43–44.), and certain industries are dominated by very small enterprises. The cultural and creative industries, and the music industry in particular, fall under these problematic areas.

The earlier statistical guidelines suggested various forms of improvements that required either special projects between national statistical authorities and representative music industry bodies or were mandated by national authorities themselves. Without these special actions, it is not possible to create similar indicators for the music industry that exists, for example, for car manufacturing, paper production or electricity distribution.

In Open Music Europe, we have no legal mandate to carry out such experiments, but based on the *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* (Open Music Europe 2023f), we will try to persuade the Slovak national authorities to carry out one; furthermore, we provide

guidelines on how the music industry's representative (collective) organisations can take action themselves and produce similar indicators. To do this, we briefly review the role of business registers, and we outline the idea of the Music Industry Business Register (MIBR.)

The Open Music Europe Data Management Plan follows the xxx in how it designs and carries out statistical data collection, and it should serve as a template for music industry organisations carrying out similar data collection on their own initiative (i.e., without the legal mandate granted by EU regulations to national statistical authorities of the member states.)

Data collection

Statistical data is collected either from accounts or surveys. Accounts, in statistical terms, are consistently collected data sources. For example, royalty accounts, which follow the general accounting principles applied in all EU countries, with industry-specific further data organisation, are the correct sources of authors' and neighbouring rights (publishing and recording copyright) revenues. Reprex and ALOADED will demonstrate the use of such accounts for statistical productions in Open Music Europe in a way other stakeholders can replicate. In this case, the royalty accounts of ALOADED are a microdata source (non-aggregated observations or measurements of characteristics of individual statistical units without direct identifiers.) The word `microdata` refers to the fact that we want to avoid the distribution of this highly confidential data specific to ALOADED and its clients, and we want to use it for a more general and aggregated data creation.

When no standard form of data registration and accounting exists, statisticians like market researchers rely on constructing surveys. The word survey originally meant a systematic examination and recording of the area and features of (an area of land) to construct a map, plan, or description. Such surveys are still used in natural sciences or cartography. Still, in business and social sciences areas, it is used metaphorically for questionnaires that are filled out about statistical units (music companies, performing music groups) either by the statistical unit itself or by the surveyor. In our experience, the best music industry surveys are administered in face-to-face interviews, when the surveyors pre-fill certain information on a questionnaire (including information on the time and place of the interview, its length, and the cooperation level of the interviewed person) and other administers the answers given orally by the surveyed statistical unit. Due to cost considerations, self-administered surveys, such as online questionnaire forms, are often used; however, in this case, the survey is filled out by a music professional and not a statistician, and misunderstandings are unavoidable.

The best way to collect statistical information is to rely on, whenever a high-quality account is available, this account itself; and use well-designed surveys with quality assurance to collect the rest of the information. In both cases, data protection rules must be adhered to at all times.

The Music Economy Data Pillar

Data needs assessment

Existing data sources

Data coordination

A Statistical register is a continuously or regularly updated set of objects for a given population. It contains information on identification and accessibility of population units as well as other attributes which support the surveying process of the population. The Statistical register should contain the current and historical statuses of the population and the causes, effects and sources of alterations in the population. Statistical register are stored in a structured database.

The role of the business register is to provide a framework for the data collection to help establish precisely the target population of the data collection and carry out the surveying and microdata linking. For example, if a nationally representative organisation like SOZA in the Slovak Republic sets up the Slovak Music Industry Business Register (SKMIBR), then this enables us to collect systematic data on the Slovak music businesses in a way that we can create similar statistics to Eurostat and the Statistical Office of Slovakia.

In previous years, the CEEMID project set up a very simple and not formalised business register for Hungary and Slovakia to create the first Hungarian national industry report and subsequent Artisjus reports, then built on this experience for surveying for the Slovak Music Industry Report. We also attempted to create a Croatian version. These experiences will be quoted for practical guidance. Statistical units

An `Observation unit` or *statistical unit* is an identifiable entity about which data can be obtained. If we create statistics on the Slovak music industry, we must observe all units that professionally contribute to the economic value creation and output of the industry in Slovakia. If the person is a natural person, such as a freelance music performer or a freelance sound engineer, then the statistical unit is also a data subject in the language of the General Data Protection Regulation (GDPR).

Regardless if we want to use questionnaires in surveys or we want to sample royalty accounts, we must be able who can answer the questions or provide us with accounting data: we need to define the geographical scope of our data collection and its economic (industry) scope.

Business registers have a "live" component: the registrar must include new industry players at all times and delete deceased entities. A good starting point for this can be the list of legal and natural persons represented by music industry organisations in a given country. SOZA, for example, collects authors' rights (copyright) revenues on the territory of Slovakia and must identify the intellectual property owners. Therefore it has an up-to-date list of composers and publishers who make their music works available on the territory of the Slovak Republic. SLOVGRAM has a similar list of entities that release recordings in Slovakia or have significant business revenue from neighbouring rights. These membership and representation lists provide a good base for the "live" component of the SKMIBR.

Because the administration of copyrights follows national jurisdictions, in this case, it is logical to create a national business registry.

Footnote: From a technical, statistical point of view, it is important to notice that this registry can collect data on a domestic view (statistical units operating on the territory of Slovakia) and national view (statistical units domiciled for tax or residency purposes in Slovakia.)

The voluntary SKMIR should follow as much as possible the actual national statistical business registry in Slovakia and provide more industry-specific detail when the national registry does not have it. In other words, it should use the same data definitions as the national authorities' business registry.

Suppose the main purpose is data collection on publishing and recording parts of the music industry. In that case, the national copyright jurisdiction may be the best geographical frame for setting up a business registry. If we want to observe the live music industry, it may be more feasible to work on the metropolitan area/provincial/state level.

All in all, a music industry business register must be able to clearly define who can and who cannot be a subject of data collection for the music industry: it must define the statistical units which are considered to be players of the music industry and whose data must be rejected, or simply not asked, because they are not part of the industry. The geographical coverage must also be defined. This defines the "business demography" of, for example, the music industry of Slovakia or Berlin.

In our vision, the SKMIBR is a voluntary registry created by the existing administrative records of membership organisations that are public by law and which allow the inclusion of other statistical units, including natural personas as data subjects, who do not wish to become members or give a mandate for representation in the maintaining organisations. However, they are still important constituents of the Slovak music industry. Maintaining the business registry needs to be very carefully designed from a data protection point of view, and the procedures for linking microdata also require strict rules that conform to GDPR in the case of natural persons and other business confidentiality and statistical data protection rules in the case of legal persons. The earlier experience of the CEEMID project is a good starting point.

Surveys can be census-like when we try to record data on all statistical units or based on a sample of the statistical units that well represent the entire population. It is important that we can move from the "live" business register to a more practical format. Suppose we want to collect annual data on the Slovak music industry. In that case, we must define a date or date range when we query the live business registry to define the population which counts into the 2023 music industry data collection. This will necessarily exclude a new music publisher starting and registering a business on the 2nd of January, 2024.

The membership lists of collective management organisations are a very good starting point because the national copyright laws of the European Union (as harmonised by the Copyright Directive) make the membership lists of such organisations public; in other creative industries, we may not find such a well-defined and public registry. The advantage

of a public registry is that we consider census-like data collection and sample-based at the same time.

For example, for the members that operate as a legal persons, other legislation makes some of their business data public by law. Because legal personality is abstracted from the concept of natural personality, it is generally assumed that an abstract legal person can be known: its name, seat, and main financial records are public, so other stakeholders wishing to contract or work with this entity can get to know the "person". In our view, it is a necessary first step to connect such data to the music industry business registry, which can be seen as a first census of the industry.

Data collection

PSI

This preliminary census will differ from the data that the Slovak national statistical authority or Eurostat can provide to us. They work with a national business registry that does not use the membership list of music organisations. Instead, it uses the so-called NACE economic activity codes of the enterprises.

Connecting NACE codes with the music industry is impossible for three reasons:

1. The music industry does not have a NACE code;
2. creative enterprises are commonly operating in several economic activity fields and may have several NACE codes;
3. National business registries do not contain a full view on very small enterprises and often do not include all freelancers, the most common working form in the music industry. These are the main reasons why we must construct the music industry business registry in the first place.

In practice, the music industry business registry must create a mapping to subsets of the NACE classes in each country.

Linking

A business registry is often a source of data itself or can provide metadata, i.e., factual statements about data that can assess the qualities of such data. Data linking is possible on both the microdata level and the statistically aggregated data level. Survey harmonisation techniques that Open Music Europe uses based on earlier CEEMID experience ensure that data collected in a survey is easily linked with either more precise microdata or at least with other statistically aggregated but more precise data sources.

It is desirable to ask the surveyed statistical units to give us legal permission to connect their administrative records with their collective management organisation or other membership organisation, or their label or distributor. For example, if we want to know the most important territories and regions where a creator's music was listened to on DSPs, it may
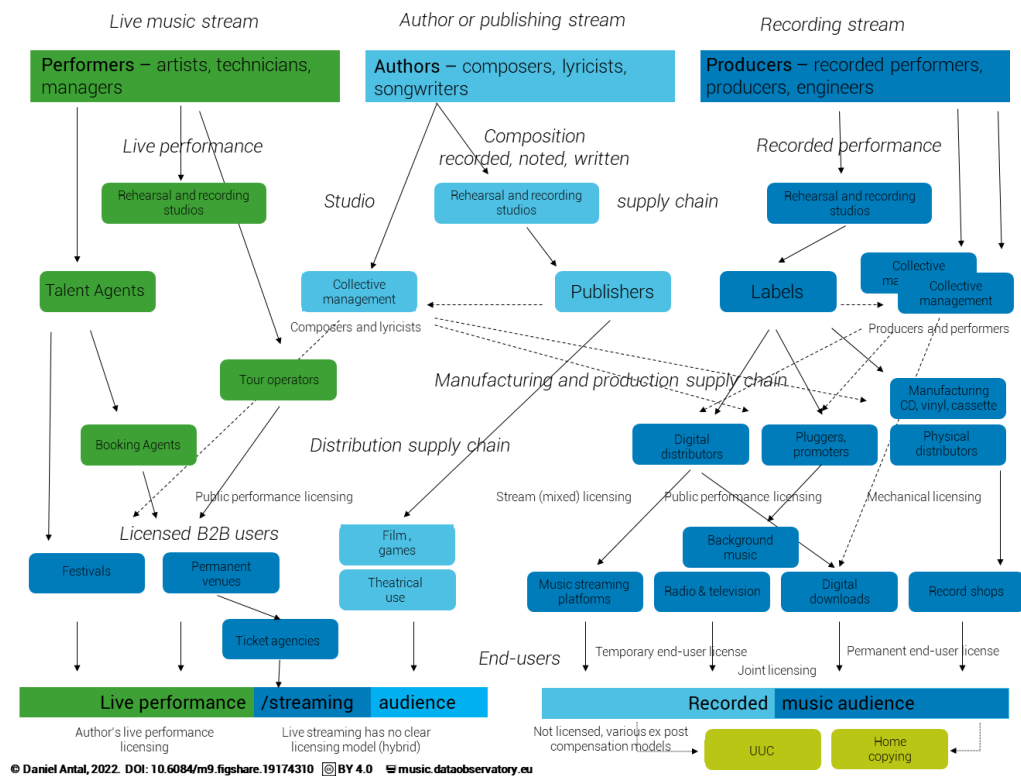
Figure 2: The music industry value chain; adopted from the model used to map the Croatian music industry.

be difficult, time-consuming or even impossible for a respondent to fill this information onto a questionnaire text box. But with access to the person's distribution list or royalty account, we can retrieve this information precisely.

Surveys

Most of the data gaps defined by the EMO Feasibility Study cannot be filled with such public data, and it is unlikely that the cost/benefit considerations of surveying would allow a census-like data collection beyond the data that these enterprises already made public. Therefore, the most important function of the SKMIBR in Slovakia is the preparation of the surveys that target only a sample of the relevant music industry preparation.

In the past years, CEEMID used a less rigid framework that considered that some music industry players are not associated with SOZA or Artisjus, for example, in Slovakia and Hungary. Our experience in the preparation of surveying the Hungarian, Slovak, and Croatian music industries first, then moving on to many other countries, was that for the music industry stakeholders, it is always a cost/benefit decision itself if they join a collective rights management organisation, even if the choice of not joining (and registering an intellectual property) results in revenues that will never be paid to the music creator or the business entity.

It is important to include as wide a range of stakeholders in the survey creation process as possible.

The way CEEMID handled these problems is that it created a meta-registry: it produced anonymous surveys that allowed data entry by music professionals who were not associated with the collective rights management organisations but who were invited by trusted members or associations. For example, in Hungary, the national DJ association and Zenészbázis, a civil society organisation that was very critical towards the Hungarian collective management organisations, agreed to distribute our surveys. We then included data points in the survey that could be statistically (but not individually) connected with Artisjus's and SOZA's data in-house.

---

i Lessons from the Central European Music Industry Report

When the CEEMID project created the *Central European Music Industry Report* (Antal 2020), we failed to create a proper collection framework for Moldva because the country had weak institutions. In Armenia, due to the general mistrust of ARMAUTHOR, we created a private registry to collect data. This is understandable: creators base their membership and representation decisions on comparing the likely annual income and the cost of administration, and they often consider subjective criteria such as the justice of royalty distribution. When CEEMID started to work in Slovakia, the Slovak collective management organisations could distribute far less income than the Hungarian organisations, and we found a higher segment of the business demography that was not associated with these organisations. In Hungary and Slovakia, we realised that hip-hop creators are often missing from the membership

---

list because they produce music that has lyrics which is too explicit for radio play or to be used as background music. As public performance revenues are the main income sources in the Central and Eastern European region, it is logical that creators who will not have such revenue do not register. In Open Music Europe, we see that Musicautor faces a more severe problem than SOZA did due to low revenue.

The *Lessons from the Central European Music Industry Report* above show that already the population definition requires an intimate knowledge of the domain. The actual data needs assessment is also based on the domain knowledge of the collective music organisations. For example, we do not want to collect data that is already available or can be created in a more comprehensive, census-like manner: connecting membership data with public company registry data is almost certainly cheaper and more reliable than surveying. When we rely on questionnaires, we must assess what can be asked in a personal survey and an enterprise survey. The questions must be set to harmonise with statistical business surveys for comparability, on the one hand, and the target population understands them. In other words, we must frame the questions so that a music professional will likely give the information the statisticians want. This is challenging because, in the music industry, most business is led by creative people who only rarely have any formal or informal training in business administration and use business information in a very informal way.

CEEMID has used questionnaire items that were standardised with the methodology created by the ESSNet-Culture working group of Eurostat and participating EU national statistical authorities (Haan and Adolfsen 2008; Bína, Vladimir et al. 2012; Haan and Broek 2012), mainly based on pre-existing best practices developed in the Netherlands. The ex ante harmonisations of questionnaire items (the same question is asked in the same format and with the same national language translations) are one of the prerequisites of joining data from surveys taken at different times, in different countries, or by different organisations. This practice has been developed by CEEMID since 2014 and further improved by the SurveyHarmonies project of SINUS and REPREX with the help of MusicAutor, SOZA, and Artisjus in 2023. It was also the basis of the development of [retroharmonize] R language software package hosted by the rOpenGov community, organised by the University of Turku Data Science group and actively developed by REPREX. Open Music Europe will continue to develop these methodologies and tools for creating a music economy, diversity surveys and collecting data on the societal and sustainability aspects of music.

For example, CEEMID pioneered the semi-open anonymous music professional survey format. To ensure that the data collected from these surveys remains representative, we included questions about royalties collected from the participating collective management organisations like SOZA in the Slovak Republic or Artisjus in Hungary. We collected survey data from a quasi-open population until we found a 95 or 99% correlation between the self-reported income of unknown music professionals from SOZA and Artisjus and the actual annual distribution of the organisations (which was provided to the researchers in an anonymous vector of numbers to establish the true distribution of the surveyed quantity.) This is a simple form of data fusion or statistical data linking (See, for example, the Slovak Music Industry Report—Správa o slovenskom hudobnom priemysle (Antal 2019).).

Reproducible exploitation of the data

Towards exploitation pathways

Linking data and microdata

References

List of Figures

Bibliography

Antal, Daniel. 2019. 'Slovak Music Industry Report [Správa o slovenskom hudobnom priemysle]'. https://doi.org/10.17605/OSF.IO/V3BE9.

———. 2020. 'Central And Eastern European Music Industry Report 2020'. CEEMID, Consolidated Independent. https://doi.org/10.13140/RG.2.2.21450.31686.

Bína, Vladimir, Chantepie, Philippe, Deboin, Valérie, Kommel, Kutt, Kotynek, Josef, and Robin, Philippe. 2012. 'ESSnet-CULTURE, European Statistical System Network on Culture. Final Report.' Edited by Frank, Guy. https://ec.europa.eu/assets/eac/culture/library/reports/ess-net-report_en.pdf.

Commission, European, Sport Directorate-General for Education Youth, Culture, M Clarke, P Vroonhof, J Snijders, A Le Gall, et al. 2020. *Feasibility Study for the Establishment of a European Music Observatory : Final Report.* Publications Office. https://doi.org/doi/10.2766/9691.

EUR-Lex. 2018. *Consolidated text: Regulation (EU) No 99/2013 of the European Parliament and of the Council of 15 January 2013 on the European statistical programme 2013-17 (Text with relevance for the EEA and for Switzerland; no longer in force).* http://data.europa.eu/eli/reg/2013/99/2018-01-01/eng.

European Parliament and the Council. 2013. *Regulation (EU) No 99/2013 of the European Parliament and of the Council of 15 January 2013 on the European statistical programme 2013-17. Text with relevance for the EEA and for Switzerland [no longer in force].* http://data.europa.eu/eli/reg/2013/99/oj/eng.

———. 2019a. *Directive (EU) 2019/1024 of the European Parliament and of the Council on Open Data and the Re-Use of Public Sector Information (Open Data Directive).* http://data.europa.eu/eli/dir/2019/1024/oj.

———. 2019b. *Regulation (EU) 2019/2152 of the European Parliament and of the Council of 27 November 2019 on European Business Statistics, Repealing 10 Legal Acts in the Field of Business Statistics (Text with EEA Relevance).* http://data.europa.eu/eli/reg/2019/2152/oj/eng.

European Statistical System Committee. 2018. *European Statistics Code of Practice — Revised Edition 2017. English Version.* Luxembourg: Publications Office of the European Union. https://doi.org/10.2785/798269.

European Statistical System (ESS). 2019. 'Quality Assurance Framework of the European Statistical System. 2019 Edition (Version 2.0)'. https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646.

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.

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

Haan, Jos de, and Anna Adolfsen. 2008. *De Virtuele Cultuurbezoeker - Publieke Belangstelling Voor Cultuurwebsites.* SCP-Publicatie 2008/9. Den Haag, the Netherlands: Sociaal en Cultureel Planbureau. https://archief18.archiefweb.eu/archives/archiefweb/20200311100055/https://www.scp.nl/dsresource?objectid=5c6903f1-e6ae-4e62-9a9b-6d5e1529756a&type=org.

Haan, Jos de, and Andries van den Broek. 2012. 'Nowadays Cultural Participation - an Update of What to Look for and Where to Look for It'. In *ESSnet-CULTURE, European Statistical System Network on Culture. Final Report.*, 397–417. Luxembourg. https://ec.europa.eu/assets/eac/culture/library/reports/ess-net-report_en.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.

Open Music Europe. 2023a. 'Open Music Europe (OpenMusE) – An Open, Scalable, Data-to-Policy Pipeline for European Music Ecosystems'. https://doi.org/10.3030/101095295.

———. 2023b. 'Repository: Report on Music Diversity and Circulation in Europe. GitHub'. Open Policy Analysis Repository with Standardized File Structure. 2023. https://github.com/dataobservatory-eu/report-music-diversity-circulation-europe/.

———. 2023c. 'Report on Music Innovation & Technology in Europe. GitHub'. Open Policy Analysis Repository with Standardized File Structure. 25 March 2023. https://github.com/dataobservatory-eu/report-music-innovation-technology.

———. 2023d. 'Repository: Report on Music, Society, and Citizenship in Europe. GitHub'. Open Policy Analysis Repository with Standardized File Structure. 25 March 2023. https://github.com/dataobservatory-eu/report-music-society-citizenship.

———. 2023e. 'Repository: Report on the European Music Economy. GitHub'. Open Policy Analysis Repository with Standardized File Structure. 26 March 2023. https://github.com/dataobservatory-eu/report-european-music-economy.

Open Music Europe, Ministerstvo kultúry SR AND. 2023f. '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/.