

Table of contents

1.1 Purpose of data generation in relation to the project (NO UPDATE ON DATASET LEVEL)	2
Datasets	3
Structural Business Statistics (OpenMusE)	4
Open Policy Analysis Guidelines	5
1.4.1 Data Summary: Music Economy	6
Size	7
1.5 Data Dissemination: Observatory	8
1.6 Data Reuse (NOT ON DATASET LEVEL)	8
2. FAIR data (NOT ON DATASET LEVEL)	10
2.1 Making data findable, including provisions for metadata	10
2.1.1 Use of persistent identifiers	10
2.1.2 Rich metadata for discovery	12
2.1.3 Search keywords	12
2.1.4 Machine actionable metadata	13
2.2 Making Data Accessible	13
2.2.1 Repositories	13
2.2.3 Metadata	14
2.3. Making Data Interoperable	15
2.3.1 Commonly used ontologies, vocabularies, and data models	16
OpenMusic / Wikibase class	17
Equivalent SDMX concept	17
Equivalent W3C Data Cube / W3C class	17
Description	17
2.3.2 Qualified references	19
2.3.3 Data exchange and reuse across disciplines	19
2.3.4 Data exchange in the music sector	20
2.4 Increase data reuse	20
2.4.1 Documentation standards	20
2.4.2 Good data semantics	21
2.4.3 API access	21
2.4.4 Data licensing	21
2.4.5 Other data	22
2.5 Quality Assurance	23
2.5.1 Eurostat indicator design principles	23
Reproducible research	23

Internal peer review	23
3. Allocation of resources (NOT APPLICABLE ON DATASET LEVEL)	24
4. Data security	24
5. Ethics	25
Data Catalogue	25
References	26

1.1 Purpose of data generation in relation to the project (NO UPDATE ON DATASET LEVEL)

The Open Music Europe Project itself can be seen as a data management project. In other words, the Data Management Plan is not only aiming at making the data generated or reused by the project further findable, accessible and reusable in an interoperable way, but data generation is a central aim of the grant. Our *Grant Agreement* and the project used the data gap analysis of the *Feasibility study for the establishment of a European Music Observatory*. This data gap analysis is dated and does not contain strictly defined data management instructions or a concise definition. We treat it as a starting point.

The three main objectives of the project are:

- **O1 identify European music data gaps;** T1.1, T2.1, T3.1 and T4.1 will suggest ways how to fill the data gaps identified before Open Music Europe and discover since and offers stakeholders consultation on them.
- **O2 bridge these gaps** by re-processing or collecting raw data and generating new data assets. T1.2, T2.2, T3.2, T4.2 tasks will acquire (collect or reuse) raw data that can form the basis of business or policy indicators or other research inputs in a processed form.
- **O3 empowers the European music ecosystem's stakeholders** to use the data to solve important business, research, or evidence-based policy problems. This is the role of T1.3, T2.3, T3.3, T4.3.
- The **Open Music Observatory will disseminate these data assets** so that they are added with new, FAIR features to the Digital Music Observatory minimum viable product (MVP.) In this regard, the projects' dissemination and data management plans are harmonised.

The subsequent subchapters of the **Data Summary** do not present the data in the functional breakup of the Grant Agreement but according to the subdomains or the “pillars” of the EMO Feasibility Study.

- [Data Summary: Music Economy](#) presents the data management of T1.1 (objective 1), T1.2 (objective 2), T1.3 (objective 3).

Apart from this Data Management Plan, or DMP, the project has a Dissemination, Communication and Exploitation Plan (DCE Plan). Because data management is not an auxiliary activity of this project, but the main project objectives are aiming at the generation of new data in terms of the dissemination and exploitation of data assets and the live policy documents (which empower the stakeholders to use the new data assets functionally), the DCE Plan and the DMP are coordinated.

The main aim of the DMP is to inform potential users of the commonly understood data in the forms of datasets, datacubes, and databases of Open Music Europe visible via metadata and, whenever possible, with data catalogues and disseminated data. As required, we also make provisions for “other data”, i.e. visualisation files, texts, blog posts, and more complex documents in lesser detail, as their access and use require less management information because their dissemination and communication are the focus of the Dissemination, Communication and Exploitation Plan. See [1.5 Data Dissemination: Open Music Observatory](#) for further details.

Purpose of data generation in relation to the project

The most relevant project actions are:

- Develop policy-relevant indicators for the total economic value of music (WP1)
- Develop new survey methods for capturing scarce data
- “Provide methodologies for capturing the economic and societal value of music”

Apart from this, as data management is a primary activity of the project, it has a project action *Develop new software for rendering fragmented, hidden, and unharmonised/unprocessed data usable* that has a supporting role in data collection and data generation.

The [Open Music Europe Data Management Plan](#) follows the *Horizon Europe Data Management Plan Template, Version 1* (OpenAire 2021).

Datasets

In statistical datasets, each observation and component—dimensions, measures, and attributes—is defined with multilingual labels and semantic mappings:

- wd:P140 (dimension) qb:DimensionProperty
- wd:P141 (measure) qb:MeasureProperty
- wd:P142 (attribute) qb:AttributeProperty
- wd:P146 (has code list) sdmx:codeList

Structural Business Statistics (OpenMusE)

The *Structural Business Statistics (OpenMusE)* ([Q600](#)) datacube is derived from the Eurostat Enterprise statistics by size class and NACE Rev. 2 activity (from 2021 onwards) datacube with filtering out relevant industries, and establishing the total number of enterprises and their total turnover.

It contains two indicators: [number of enterprises](#) and [net turnover of enterprises in million euros](#), aggregated according to two approximate definitions of the music industry. The *CI-CERONE* flag marks observations that fall within the minimum or expanded NACE ranges defined in the CICERONE project, while the other (the *OpenMusE SBS* flag) marks observations belonging to the core or extended music-sector definitions developed in OpenMusE. These flags enable consistent extraction, comparison, and modelling of music-related economic activity across otherwise incompatible datasets.

Note

This part is for illustration only, and does not need repetition in the DMP.

The dataset [Q600.ttl](#) (also available as [Q600.nt](#), [Q600.jsonld](#), [Q600.json](#), and [Q600.rdf](#) in XML) represents a machine-readable model of the Eurostat table “*Structural Business Statistics (OpenMusE)*”. It was reconstructed in the OpenMusic Wikibase <https://reprexbase.eu/openmusic/>) to demonstrate alignment between European statistical metadata (SDMX) and linked-data publication formats used in the cultural-heritage domain.

The dataset uses RDF/Turtle format and aligns Wikibase schema elements ([wikibase:Item](#), [wikibase:Property](#)) with W3C and SDMX standards. Dataset-level metadata ([schema:Dataset](#)) record provenance ([prov:wasDerivedFrom](#), [dct:source](#)) and licensing (CC0 1.0).

These alignments allow full interoperability between Eurostat’s SDMX-based dissemination and linked-data environments such as the EU Open Data Portal.

The dimensions of the dataset:

- [NACE economic activity \(dimension\)](#)
- [time \(dataset dimension\)](#) *currently not added, but it would be absolutely essential*
- [geopolitical entity \(dimension\)](#) *not present now but would be absolutely necessary*

The measured variables of the datacube:

- [number of enterprises](#)
- [net turnover of enterprises in million euros](#),

The attributes of the dataset: *These are currently not recorded, but it would be a very good practice, and they are present in the data that you retrieve from Eurostat,*

- [unit of measure \(attribute\)](#) (not present now)
- [frequency \(attribute\)](#) (not present now)
- [OpenMusE SBS music-sector flag](#)
- [CICERONE music classification flag](#)

The dataset is openly available for reuse and serves as a model for FAIR statistical data integration in the cultural and creative domains.

Purpose and context

! Economics summary

You should describe here in one paragraph what this indicator can be used for and what not. It would need a word of caution: this is created with retroactive filtering, not with collecting the raw data from music organisations.

Data types and formats

The metadata are provided as **RDF triples** (Turtle .ttl), optionally serialised as JSON-LD, N-Triples, and RDF/XML. All identifiers resolve to dereferenceable HTTP URIs under <https://reprexbase.eu/openmusic/Special:EntityData/>, for example: <https://reprexbase.eu/openmusic/Special:EntityData/Q600>. Each dataset entity includes an access point (`om:P130, has access point`) linking to a DOI that resolves to the corresponding Eurostat dataset page, where users can download subsets of the data cube in multiple formats.

Open Policy Analysis Guidelines

Our Consortium pledged to follow the Open Policy Guidelines (Level 3). (BITSS 2019; Hoces de la Guardia, Grant, and Miguel 2020). Within the Horizon Europe research framework programme, some open science practices are mandatory for all beneficiaries per the grant agreement, including using the FAIR principles (European Commission 2023, p43). Compliance with the FAIR principles is laid out in Parts 2-5.

According to the Horizon Europe Programme Guidelines, recommended practices “... are open science practices beyond the mandatory ones, such as involving all relevant knowledge actors, including citizens, early and open sharing of research, output management beyond research data, open peer-review. This is a non-exhaustive list of practices that proposers are expected to adopt when possible and appropriate for their projects. Finally, certain work programme topics or call conditions may encourage specific additional open science practices.”

In *Open Music Europe*, we chose to comply with such an additional standard, the *Open Policy Analysis Guidelines* (Hoces de la Guardia, Grant, and Miguel 2020), which are a practical standard intended to one of our core target groups: cultural policymakers.

Open Policy Analysis Guidelines

- OPA 4: Share raw (or analytic) data and materials in a way that the analysis is reproducible with minimal effort. Level 3: Analytic and raw data are made available through a trusted repository. Detailed instructions are provided for accessing raw data that is proprietary or contains sensitive information.
- OPA 6: Standardise the file structure so that materials are organized in a way that is accessible to an informed reader. All project components are organized in a self-contained folder using a Standard File Structure (SFS), and a readme file is included.
- OPA 9: Level 3: All team members use version control software and track changes in a shared project repository.

The [wp1-to-music-observatory](#) contains the reproducible code and the raw values of the tables for the *Structural Business Statistics (OpenMuse)* ([Q600](#)) [... continue here] datacube following the standards of the R statistical environment and language.

The **data-raw** folder contains downloaded raw data.

The **data** folder contains publication ready data. The **documentation** folder contains the documentation.

The **bib** folder contains the bibliographical files (original data sources, metadata definitions, and reports and publications that use the data.)

The **R** folder contains R scripts to produce the data tables.

This information can be found on the README page of the repository folder.

1.4.1 Data Summary: Music Economy

Data Generated & Reused: Music Economy

The expected outcome of **Objective 1 MAP the policy and data landscape** is to “develop policy-relevant indicators for the total economic value of music.” The indicator development guidelines of Eurostat (see *2.5 Quality Assurance*) will be used for the general methodology.

WP1 will provide indicators capable of capturing the full economic value of music, including zero-price uses. For this, existing data gaps are identified, analysed and filled. i.e., currently “hidden” data will be reported, MSMEs that currently do not provide data are integrated, zero-price use is measured, etc. Furthermore, methods to effectively act on this data are demonstrated. A pilot study demonstrates exemplary how improved data

collection on the value of music can be leveraged to improve artist revenues. Examples of potential indicator candidates developed in WP1 include (but are not limited to): Employment; the value of EU's music sector; structure of the market; impact of the not-for-profit sector on the overall economy of the music sector; Neighbouring rights; Music publishing; Independent music companies; Live music; Export music; Music retail or in-store public performance; Financing of the music sector; Live music regulation; Copyright regulations and evolution of copyright regimes; etc.

WP1 will provide indicators capable of capturing the full economic value of music, including zero-price uses. For this, existing data gaps are identified, analysed and filled. i.e., currently "hidden" data will be reported, MSMEs that currently do not provide data are integrated, zero-price use is measured, etc. Furthermore, methods to effectively act on this data are demonstrated. A pilot study demonstrates exemplary how improved data collection on the value of music can be leveraged to improve artist revenues.

Examples of potential indicator candidates developed in WP1 include (but are not limited to): Employment; Value of EU's music sector; Structure of the market; The impact of the not-for-profit sector on the overall economy of the music sector; Neighbouring rights; Music publishing; Independent music companies; Live music; Export music; Music retail or in-store public performance; Financing of the music sector; Live music regulation; Copyright regulations and evolution of copyright regimes; etc.

Reuse & collection

The dataset is part of the [Music Economy Pillar](#) of the Open Music Observatory.

Data generation

The data is generated by the `TABLE_2_3_4_REPLICATION.R` script that can be found in the R library of the OPA folder.

Size

`t3_by_nace` contains 24x6 cells of data with an approximate size of 0.9 KiB in CSV serialisation.

1.5 Data Dissemination: Observatory

! Dissemination plan

Once the table is finalised, it will be placed on the Open Music Observatory interactive browser, the EU Open Data Portal and Zenodo.

1.6 Data Reuse (NOT ON DATASET LEVEL)

This is a summary table; for further details, see the following chapters of the DMP.

! Important

Most the the data below needs no update, only the OPA folders need to be linked. This is not a dataset but DMP level

Data types	Collecting	Storage	Access
survey mirco-data datasets	Collected by SINUS and REPREX. Reused under the Open Data Directive.	Sensitive data is stored by collecting entity. Statistical microdata on GitHub repositories (temporary) and Zenodo (long-term)	No access to sensitive microdata. Non-identifiable datasets will be free to reuse with open data license. reused data will be made further reusable on similar terms we receive them.
administrative records	MADE, LOADED, ARTISJUS, SOZA, MUSICAUTOR and third party administrative records. Processed by Consortium members.	Sensitive data is stored by collecting entity. Non-sensitive and limited subsets from M17 in Digital Music Observatory as linked open data.	No access to sensitive microdata. Limited access to microdata with data linking. Linking is possible with individual licenses.

Data types	Collecting	Storage	Access
statistical and indicator datasets	Reused from various open government sources. Processed from Consortium member microdata by Consortium.	Temporary access on GitHub (pre-processed and not human-controlled.) Open Music Observatory on Zenodo From M17 in Digital Music Observatory API.	Released as open data.
data visualisations	Created by members of the Consortium or Consortium.	FigShare open science repository.	CC-BY license.
software code	UTU, REPREX in open collaboration with qualifying volunteers.	GitHub (development versions) Comprehensive R Archive Network (CRAN) peer-reviewed releases. Zenodo (development and peer-reviewed releases.) GitHub	Free access, with open source license statement for each software.
statistical processing code	UTU, REPREX in collaboration with microdata owners.		Whenever possible, released as software. Redactions are possible for data security purposes.
documents	Open Music Europe Consortium and its members	GitHub (see references) Final versions on Zenodo (and in journals)	Free access for non-commercial use.
blogpost documents	Open Music Europe Consortium and its members	This is the example Link to your blogpost, add to Bib file.	CC-BY license
live policy documents	Open Music Europe Consortium and its members	Digital Music Observatory web resource	Complex document, parts (data, visualisation, text, code) under various CC licenses and open source licenses.

Data types	Collecting	Storage	Access
metadata	Open Music Europe Consortium and its members	See details of DMP.	CC0

2. FAIR data (NOT ON DATASET LEVEL)

2.1 Making data findable, including provisions for metadata

Making data findable

- Will data be identified by a persistent identifier?
- Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.
- Will metadata be offered in such a way that it can be harvested and indexed?

2.1.1 Use of persistent identifiers

Making data findable

- Will data be identified by a persistent identifier?

Our data collection, processing and incremental addition to our datasets will be a daily activity, and we will use temporary (buffer) and persistent identifiers for all data assets.

ORCiD: ORCID provides a persistent digital identifier (an ORCID iD) that you own and control, and that distinguishes you from every other researcher. We use it to identify contributors to our data assets. *Arianna has an ORCiD, it was used to identify her*

DOI: All our datasets will have a versioned DOI on Zenodo; high-frequency data will have temporary identifiers and will only be periodically released on Zenodo with a new DOI. All our visualizations will have a DOI on FigShare, which is a global open repository particularly designed to make visualizations reusable. When datasets are visualised, then the visualization

will be connected to the persistent identifier (DOI) of the dataset. *Once published the dataset will have a DOI*

VIAF: VIAF explores virtually combining the name authority files of national-level authority files into a single name authority service. We use the VIAF ID as a PID. We use it in the creation of diversity indicators, in the original inventory microdata datasets. *Only if Arianna has a VIAF cluster ID*

Datasets

Visualizations

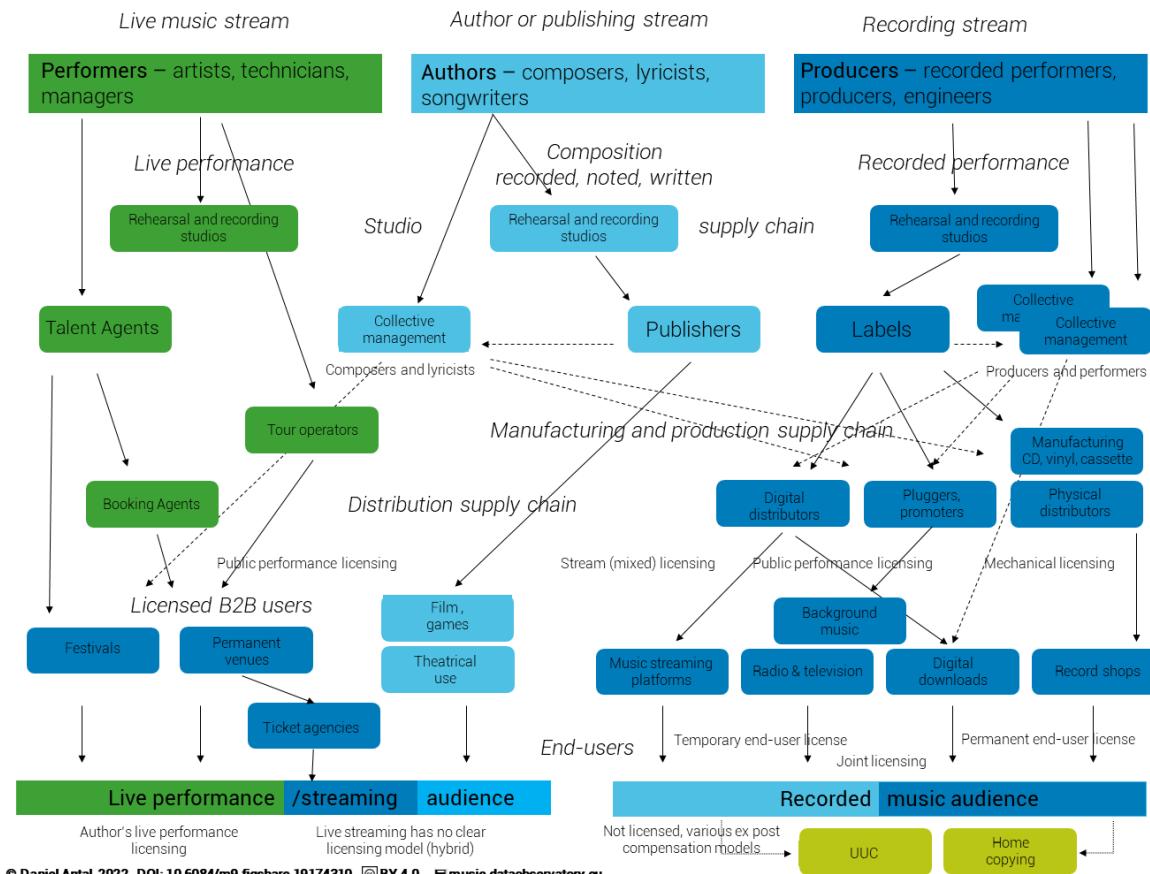


Figure 1: PID for Visualizations: All our visualisations will have a DOI on FigShare

Visualisations

Add info about the visualisations of datasets (they can be uploaded to FigShare or Zenodo, as with journal article figures)

Documents

Live policy documents: Live policy documents, as their name suggests, are live updating, often very frequently. Similarly to their (embedded) high-frequency datasets, they will be saved periodically and released on Zenodo with a versioned DOI.

Static documents: Under the current European legislation, all informative files, including text documents fall under the legal definition of data. Their dissemination is defined in the Open Music Europe consortium's Dissemination, Communication and Exploitation Plan. The DMP relates to the human and machine-actionable release of more structured information that is commonly referred to as 'data'. To make our data more reusable, all data used in the disseminated static documents will be made available in human and machine-actionable dataset forms and data visualisations.

2.1.2 Rich metadata for discovery

Making data findable

- Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

All our data will be provided with the mandatory metadata of the *DataCite 4.4* standard (DataCite Metadata Working Group 2021), and in almost all cases, all recommended metadata, too.

2.1.3 Search keywords

Note

- Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

We use the controlled libraries that are supported by Zenodo.

2.1.4 Machine actionable metadata

 Note

- Will metadata be offered in such a way that it can be harvested and indexed?

Our data management practice is machine-actionable, this means that we will offer metadata in a way that it can be machine-read, harvested and indexed.

2.2 Making Data Accessible

Open Music Europe, as the name of the project suggests, wants to make as much data openly accessible for music businesses, music-related and evidence-based cultural policy and music research as possible. Most of our data in the form of datasets, and other data (i.e. visualizations, texts, complex documents and live policy documents) will be openly accessible unless otherwise stated in this DMP.

2.2.1 Repositories

 Note

- Will the data be deposited in a trusted repository?
- Have you explored appropriate arrangements with the identified repository where your data will be deposited?
- Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Our main, long-term repository for data is Zenodo. The project's 5.1 deliverable, the Open Music Observatory, had a minimum viable product prototype, the Digital Music Observatory, which has had a community on Zenodo since 2021. (Digital Music Observatory 2021b)

Our short-term repository, following the Open Policy Analysis Guidelines, provides access and reviewability to our data and inputs from the very early stage (in some cases, regarding documentation, from the proposal phase.) These repositories are stored on GitHub, and their addresses are available in the reference list of this DMP and via our project website.

Zenodo and GitHub offer continuous integration with each other. GitHub will be used as a working repository to comply with the OPA Level 3 standard, and data assets that are approved for dissemination and communication will be placed onto Zenodo.

FigShare is a repository aimed at reusable visualisations. While we can also place visualisations on GitHub and Zenodo, FigShare has a visualization-centric global audience. When

visualisations can be reused individually, we will place them FigShare, and use FigShare's DOIs as persistent identifiers. Naturally, when the primary DOI and dissemination point of a dataset and its methodological description is Zenodo and the visualisations' is FigShare, we will cross-reference these assets. The project's 5.1 deliverable, the Open Music Observatory, had a minimum viable product prototype, the Digital Music Observatory, which has had a community on Zenodo since 2021. (Digital Music Observatory 2021a)

2.2.3 Metadata

Metadata

- Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?
- How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?
- Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

The metadata of Open Music Europe data will be openly available and in human-readable, reusable, and machine-readable format(s).

In the first phase of the project (M1-M16), this will be achieved by periodically releasing all data on the Zenodo open science repository in the Digital Music Observatory collection. All data deposited and released in the repository will have metadata available in the following metadata formats: MARCXML, Dublin Core (according to OpenAIRE Guidelines), DataCite, DCAT, and JSON-LD (Schema.org).

In the second phase of the project, we will start delivering the D5.1 Open Data Observatory, which will provide API access to our data (via SQL queries and in machine-readable JSON format) and provide further machine-readable data catalogues along the aforementioned solution.

The metadata (and the data) will be deposited on Zenodo, which provides access to the data for the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.

According to the *Exploitation Plan* of the Open Music Europe Consortium, our plan is to make our Open Music Observatory, as part of a larger Digital Music Observatory, eventually part of the planned European Music Observatory. If this is not feasible, we would like to find an exploitation path that ensures the continuity of the data collection, processing, and dissemination of the project. If successful, then we will not only provide long-term access to the data and the metadata but continue to renew our datasets.

Almost all data will be provided in a way that it can be used and reused in numerous widely used open sources and licensed software applications. To enhance interoperability, we will always provide data CSV files that conform to the W3C interoperability standard about the release of CSV files with metadata.

After consulting stakeholders in the target group of our project, we may provide data in different file formats that is more comfortable for our users.

One of our projects aims is to provide a fully reproducible data-to-policy pipeline and full reproducibility in the R statistical environment and language. The R environment is both open-source and interoperable; it is available on a wide range of operational systems (i.e., various Linux or BSD distributions, including MacOS and various Windows operating systems.) Apart from using the resilient but inefficient CSV format, we will release data in native rdf formats. One of our aims in our Work Package 4 is to create an extension package to the R language (library ecosystem) that ensures an easy reproduction of our data-to-policy pipeline.

2.3. Making Data Interoperable

Note

- What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and reuse within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?
- In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?
- Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

As data interoperability intersects with data reuse, we explain some of our approach in 2.4.

2.3.1 Commonly used ontologies, vocabularies, and data models

Our datasets follow the vocabularies and definitions of the data model of the Statistical Data and Metadata Exchange (SDMX), i.e., the [SDMX 3.0 Technical Specifications](#). Section 2 Information Model: UML Conceptual Design. Version 3.0 which provides the widest possible interoperability with the data products of national statistical agencies and international agencies such as Eurostat or OECD (SDMX 2021). We will also comply with the W3C's [RDF Data Cube Vocabulary](#), which applies the datacube and dataset model of SDMX to the word wide web (W3C 2014).

The combination of SDMX and RDF Data Cube standards ensures that OpenMusic data can be automatically harvested, linked, and reused by Eurostat-compliant systems, the EU Open Data Portal, OpenAIRE and other linked open data repositories, and cultural heritage knowledge graphs such as Wikidata, Europeana and the European Collaborative Cloud for Cultural Heritage (ECCCH).

Property alignment between OpenMusic Wikibase and SDMX / RDF Data Cube

OpenMusic property	Function / role	RDF Data Cube	SDMX equivalent	equiva- lent	Example entity / comment
om:P140 (“dimension (dataset)’)	Identifies one axis of observation (e.g., time, sex, education)	sdmx:Dimension	sdmx:Dimension	sdmx:Dimension	P456 “time (dimension)” , wd:Q472 “time (dimension)”
om:P141 (“measure (dataset)’)	Quantitative variable observed	sdmx:Measure	sdmx:Measure	sdmx:Measure	P457 “persons practising artistic activities (measure)”
om:P142 (“attribute (dataset)’)	Qualifying metadata (unit, frequency, status)	sdmx:Attribute	sdmx:Attribute	sdmx:Attribute	P460 “unit of measure” , wd:Q470 “frequency”
om:P146 (“has code list”)	Links a dimension or concept to its controlled vocabulary	sdmx:codeList	sdmx:codeList	sdmx:codeList	wd:Q463 “CL_SEX”
om:P145 (“publisher”)	Identifies data provider	sdmx:DataProvider	sdmx:DataProvider	sdmx:DataProvider	Q449 “Eurostat”
om:P95 (“DOI”)	Persistent identifier	sdmx:Identifier	sdmx:Identifier	sdmx:Identifier	P2908 / ILC_SCP07

OpenMusic property	Function / role	SDMX equivalent	RDF Data Cube equivalent	Example entity / comment
om:P130 (“has access point”)	Access URL for dataset	—	dcat:accessURL / schema:url	https://doi.org/10.2908/ILC SCP01
prov:wasDerivedFrom	Provenance record	sdmx:source	prov:wasDerivedFrom	Eurostat dataset source
cc:license	Rights and licensing	sdmx:DataSetLicense (conceptual)	sdmx:DataSetLicense (conceptual)	ECO 1.0 Public Domain Dedication

Class alignment between OpenMusic Wikibase, SDMX, and W3C RDF Data Cube

OpenMusic / Wikibase class	Equivalent SDMX concept	Equivalent W3C class	Description
om:Q285 (“data set”)	concept	DataCube / W3C	Defines the structure of dimensions, measures, and attributes within the sdmx:DataStructureDefinition class
om:Q446 (“Persons practising artistic activities by sex, age, education and frequency”)	dataset	qb:DataStructureDefinition	qb:DataStructureDefinition
om:Q449 (“Eurostat”)	datasetProvider	qb:DataSet	Represents the Eurostat dataset reconstructed as RDF.
om:Q456 (“sex (dimension)”)	dimension	qb:Dimension	qb:Dimension
	variable	qb:Property	variable defining the sex of the observed population.

om:Q467 (“geopolitical entity (dimension)”)	sdmxdim:REF_DIMENSION Property variable defining the geographic area (country or region).	
om:Q472 (“time (dataset dimension)”)	sdmxdim:TIME_PERIOD Property variable defining the reference time period.	
om:Q466 (“educational attainment (dimension)”)	sdmxdim:EDUCATION Property representing the highest level of education attained.	
om:Q471 (“persons practising artistic activities (measure)”)	sdmx:Measure Property variable representing the observed value (number or percentage of persons).	
om:Q469 (“unit of measure (attribute)”)	sdmx:UNIT Property specifying the unit of measurement (e.g., percentage of persons).	
om:Q470 (“frequency (attribute)”)	sdmxdim:REQ_ATTRIBUTE Property describing the observation’s reporting frequency (annual, quarterly, etc.).	
om:Q463 (“CL_SEX”)	sdmx:CodeList Property vocabulary providing coded values for the “sex” dimension.	
om:Q604 net turnover (million euro)	ESTAT:INDIC_SBS(10.0)	Eurostat SBS 4.1 indicator
om:Q605 enterprise number	ESTAT:INDIC_SBS(10.0)	Eurostat SBS 4.1 indicator
om:Q602 OpenMusE SBS music-sector flag	OpenMusE defined	D1.2 reference

om:Q601 CICERONE music classification flag	OpenMusE defined	D1.2 reference
--	------------------	----------------

In WP1 we will use, beyond the SDMX taxonomies and ontologies the new ESCO 2.0 ontology (De Smedt et al. 2024).

These definitions were designed for statistical datasets, and they are directly applicable to our **indicator datasets**. However, with simplifications, they can also apply to our microdata, except that these datasets contain statistically not yet processed raw data. This means that the microdata datasets will contain the measures and dimensions as grouping variables, which will become the measures and dimensions of the indicator dataset after statistical aggregation.

In T3.2, we will aim to apply more and more of the DDI Standards, starting with the **DDI-Codebook 2.5** standard (DDI Alliance 2012). Later versions of this DMP will guide on the introduction of further DDI standards.

2.3.2 Qualified references

 Note

This part is basically a rewording of the requirements with the exception of handling the issue of ISWC that needs to be clarified with CISAC.

If the data set builds on another data set, if additional data sets are needed to complete the data, or if complementary information is stored in a different data set, this needs to be specified. This will be the case in many of our datasets.

From a data documentation point of view, a scientific link between the data sets needs will be described, and all data sets will be cited with the inclusion of their persistent identifiers.

This approach will be taken with our indicator datasets, which follow as closely as possible SDMX practices; complex indicators will often use as components of pre-existing datasets disseminated following the SDMX standards.

2.3.3 Data exchange and reuse across disciplines

We mainly foster interdisciplinary reuse of our indicator datasets by applying the cross-domain codebooks for data attributes and dimensions of SDMX, which are precisely designed for cross-domain interoperability.

2.3.4 Data exchange in the music sector

Not applicable here.

2.4 Increase data reuse

The Open Music Europe project aims to establish a reproducible and interoperable model for **reusable data in the European music and cultural sector**. Because data generation is a core objective of the project—not a secondary activity—our data management strategy combines open science principles with careful handling of intellectual property, privacy, and cultural rights. We apply a **tiered openness model**, ensuring that each dataset and its accompanying metadata are reusable to the maximum extent permitted by law, ethics, and partner agreements.

Our practices improve both **machine-actionable** and **human** reuse.

All data and documentation follow the FAIR principles and the Open Policy Analysis (OPA) Guidelines Level 3, ensuring full reproducibility and transparent provenance.

2.4.1 Documentation standards

Note

- How will you provide documentation needed to validate data analysis and facilitate data reuse (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?
- Will the provenance of the data be thoroughly documented using the appropriate standards?

All datasets are produced in compliance with the **OPA Guidelines (Level 3)**, which require transparent documentation and open computational workflows.

Each dataset is accompanied by a *README* and structured according to the **Standard File Structure (SFS)** used throughout the Consortium. These folders, together with Quarto or R Markdown notebooks, ensure that each data transformation is reproducible.

Provenance is recorded according to the relevant domain standards:

- **Statistical data** use SDMX and W3C RDF Data Cube provenance elements (`prov:wasDerivedFrom`, `dct:source`);
- **Survey data** follow the DDI standard for variable- and respondent-level documentation;

In the second phase of the project, the deliverable **D5.1 Open Music Observatory** will provide a consolidated, machine-actionable data catalogue (DCAT 2.0), and this DMP will be updated accordingly.

2.4.2 Good data semantics

Data are structured so that their format reflects their analytical meaning, making them easy to reuse both by humans and by software systems.

All tabular data follow the **tidy data principle**, a statistical restatement of Codd's third normal form.

The structure conforms to SDMX and RDF Data Cube definitions: columns represent **measures**, **dimensions**, and **attributes**, while each row represents one **observation** (`qb:Observation`).

Column names follow a standardised `snake_case` convention for readability and interoperability.

Observations and entities are uniquely identified with **persistent identifiers (PIPs)**, typically HTTP URIs resolving to the OpenMusic Wikibase (<https://reprexbase.eu/openmusic/>). This ensures semantic and programmatic alignment across linked datasets.

For visualisations, semantic integrity means using accurate, non-misleading forms of representation.

For textual and complex documents, it means maintaining a clear structure with summaries, references, and provenance metadata.

2.4.3 API access

We provide various level of API access to the datasets.

2.4.4 Data licensing

Note

- Will your data be made freely available in the public domain to permit the widest re-use possible?
- Will your data be licensed using standard reuse licences, in line with the obligations set out in the Grant Agreement?
- Will the data produced in the project be usable by third parties, in particular after

the end of the project?

Licensing follows a **tiered model** that distinguishes between metadata, open data, and restricted data:

- **Metadata** — All metadata describing datasets, variables, and provenance are released under [CC0 1.0 Public Domain Dedication](#).

This allows unrestricted reuse, harvesting, and aggregation in catalogues such as Zenodo, OpenAIRE, and the ECCCH data space.

- **Open data** — Aggregated or derived datasets produced by the project (e.g. statistical indicators, anonymised survey summaries, or cultural statistics) are released under [CC BY 4.0](#), unless specific partner or source conditions require another licence.

Each dataset's licence is declared explicitly in its metadata (`dct:license`) under the [om:P148](#) property.

As a general principle, the project applies an “open by default, protected by exception” policy. This ensures compliance with the Horizon Europe open-science obligations while respecting legal and ethical constraints.

2.4.5 Other data

Note

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

In addition to datasets, the project produces **software code, methodological notebooks, documentation, and live policy reports**.

These are versioned and released under appropriate open-source or Creative Commons licences (typically MIT, GPL-3.0, or CC BY 4.0).

Each output is linked to its underlying dataset and provenance record in the Open Music Observatory catalogue.

2.5 Quality Assurance

Note

Describe all relevant data quality assurance processes.

2.5.1 Eurostat indicator design principles

In our projects, we follow the best practices of key business information, statistical, and evidence-based policy indicator design. In doing so, we would like to find synergies among various recent innovations in statistics and open science. Throughout the project, we will follow the Eurostat guidelines on creating new indicators (Eurostat 2014, 2017; Kotzeva et al. 2017), which will ensure broad consensus forming among stakeholders around the objectives and methodology of the improved measurements.

Reproducible research

We follow the principles of reproducible research, which increases data quality by using open algorithms, provisioning complete data (lifecycle) history, unit testing, and facilitating both internal and external review and audit. Most of our analysis, hypotheses, software code used for data processing, and the data itself is open. Our critical software elements in the data-to-policy pipeline are regularly peer-reviewed on *The Comprehensive R Archive Network*, and we send from time-to-time their methodologies for independent scientific peer review.

Internal peer review

Internal peer review short description

Just add here a few words.

3. Allocation of resources (NOT APPLICABLE ON DATASET LEVEL)

i Note

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe

i Note

Who will be responsible for data management in your project?

i Note

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value)

Zenodo is hosted by CERN, which has existed since 1954 and currently has an experimental programme defined for the next 20+ years. CERN is a memory institution for High Energy Physics and is renowned for its pioneering work in Open Access. Organisationally, Zenodo is embedded in the IT Department, Collaboration Devices and Applications Group, Digital Repositories Section (IT-CDA-DR).

4. Data security

i Note

- What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?
- Will the data be safely stored in trusted repositories for long term preservation and curation?

The final, processed data will be synchronized between GitHub and Zenodo, or a permanent, long-term repository monthly. Our datasets are periodically updated on Zenodo (under a versioned DOI) which is planned to operate at least for the next 20+ years. Zenodo itself is funded by OpenAIRE and offers continuous integration with GitHub.

5. Ethics

Note

- Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
- Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

Consortium members and their subcontractors will be involved in creating survey metadata datasets and follow the ICC/ESOMAR International Code on Market, Opinion and Social Research and Data Analytics sets the standard of ethical and professional conduct for the global data, research and insights community. Furthermore, Consortium members, when conducting fieldwork for surveys (directly or via subcontractors), must comply with GDPR and carry out a *Data protection impact assessment* when defined necessary by the regulation.

In the case of personal surveys, the rules of GDPR and ICC/ESOMAR will be used to prevent the leakage of personal data into the survey microdata sets, which will be available for analysis. In the next processing step, indicator datasets will be created, which are statistically aggregated from a significant number of responses so that they cannot be disaggregated to guess, estimate, or predict individual personal data.

In the case of enterprise surveys, the survey is designed in a way that applies only to legal persons. In this case, the provisions similar to the *2019/1700 EU Regulation on establishing a common framework for European statistics relating to persons and households, based on data at individual level collected from samples*, considering, of course, the difference that we do not apply the rules as a statistical authority and have a different legal basis for data collection.

Data Catalogue

DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. This document defines the schema and provides examples for its use.

The data catalogue will contain human-readable tables and machine-readable (DCAT Version 2 standard) format information about the datasets created by ‘Open Music Europe’. See an example from the standard:

Properties	Example values
dct:title	“Open Music Europe Example Dataaset”@en
dcat:keyword	“music”@en, “musique”@fr, “payments”@en

Properties	Example values
dct:creator	Jane Doe
dct:issted	“2021-06-01”^^xsd:date
dct:temporal	http://reference.data.gov.uk/id/quarter/2006-Q1
dcat:temporalResolution	“P1D”^^xsd:duration
dct:spatial	http://sws.geonames.org/6695072/
dct:publisher	open-music-europe
dct:language	http://id.loc.gov/vocabulary/iso639-1/en
dct:accrualPeriodicity	http://purl.org/linked-data/sdmx/2009/code#freq-A
dcat:distribution	dataset-001-csv

dataset-001 a dcat:Dataset

Q600: *Structural Business Statistics (OpenMuse)* (Martinelli 2025)

References

- BITSS. 2019. “Guidelines for Open Policy Analysis.” Berkeley Initiative for Transparency in the Social Sciences. <http://www.bitss.org/wp-content/uploads/2019/03/OPA-Guidelines.pdf>.
- DataCite Metadata Working Group. 2021. “DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. [Version 4.4.]” DataCite e.V. <https://doi.org/10.14454/3w3z-sa82>.
- DDI Alliance. 2012. “DDI-Codebook 2.5 | Data Documentation Initiative.” DDI Alliance. <https://ddialliance.org/Specification/DDI-Codebook/2.5/>.
- De Smedt, Johan, Agis Papantoniou, Bernard Vatan, Karel Kremer, Laurent Bégin, Martin Le Vrang, Massimiliano Molinari, Niels Vandekeybus, and Pieter Fannes. 2024. “The ESCO Ontology: European Skills, Competences, Qualifications and Occupations Ontology.” Ontology specification available at <http://data.europa.eu/esco/model/2.0.0>; European Commission, Directorate-General for Employment, Social Affairs; Inclusion (DG-EMPL). <http://data.europa.eu/esco/model/2.0.0>.
- Digital Music Observatory. 2021a. “Digital Music Observatory Collection on Figshare.” Digital Music Observatory. <https://figshare.com/account/home#/collections/5784023>.
- . 2021b. “Digital Music Observatory Community on Zenodo.” Digital Music Observatory. https://zenodo.org/communities/music_observatory/.
- European Commission. 2023. “Horizon Europe (HORIZON) Programme Guide. Version 3.0.” European Commission. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf.

- 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>.
- Hoces de la Guardia, Fernando, Sean Grant, and Edward Miguel. 2020. “A framework for open policy analysis.” *Science and Public Policy* 48 (2): 154–63. <https://doi.org/10.1093/scipol/scaa067>.
- 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>.
- Martinelli, Arianna. 2025. “Structural Business Statistics (OpenMuse).” The Comprehensive R Archive Network. <https://reprexbase.eu/openmusic/Item:Q600>.
- OpenAire. 2021. “Horizon Europe Data Management Plan Template, Version 1.” https://www.openaire.eu/images/Guides/HORIZON_EUROPE_Data-Management-Plan-Template.pdf.
- SDMX. 2021. “SDMX 3.0 Technical Specifications. Section 2 Information Model: UML Conceptual Design. Version 3.0.” Statistical Data; Metadata eXchange. https://sdmx.org/wp-content/uploads/SDMX_3-0-0_SECTION_2_FINAL-1_0.pdf.
- W3C. 2014. “RDF Data Cube Vocabulary.” Edited by Richard Cyganiak and Dave Reynolds. W3C. <https://www.w3.org/TR/vocab-data-cube/>.