Music360

A 360 DEGREES PERSPECTIVE ON THE VALUE OF MUSIC

D8.2 Data Management Plan

V 1.0 August 2023

Disclaimer

The Music360 project has received funding from the European Union’s Horizon Europe research and innovation action under grant agreement number 101094872.

The opinions expressed in this document reflects only the author`s view and in no way reflect the European Commission’s opinions. The European Commission is not responsible for any use that may be made of the information it contains.

Version history

Ver.	Date	Comments/Changes	Author/Reviewer
0.1	20 April 2023	First draft	Anna Bon/Mekky Zaidi
0.2	7 July 2023	Formatted draft, interview BMAT added	Anna/ Denis/Gonçal
0.3	30 August	Ready to present at the Valencia meeting	Anna/Mekky
0.4	12 September	information BUMA added	Anna Bon and Frank Lucassen
0.5	13 September	Review round by	Conrado Carrascosa and Ingmar Leijen
1.0	14 September	Information BMAT added ; Final version	Anna Bon

Project Acronym	**Music360**	
Project Title	360 Degrees Perspective on the Value of Music	
Project Number	**101094872**	
Instrument	Research and Innovation Action (RIA)	
Topic	HORIZON-CL2-2022-HERITAGE-01-05	
Project Start Date	01/03/2023	
Project Duration	36 months	
Work Package	WP8	
Task	**T8.2**	
Deliverable	D22	
Due Date	31 August 2023	
Submission Date	15 September 2023	
Dissemination Level[1]	Public	
Deliverable Responsible	VU Amsterdam	
Version	V 1.0	
Status	Published	
Author(s)	Anna Bon, Mekky Zaidi	
Reviewer(s)	Conrado Carrascosa, Ingmar Leijen	

[1] PU= Public, CO=Confidential, only for members of the consortium (including the Commission Services), CL=Classified, as referred to in Commission Decision 2001/844/EC

Disclaimer/Acknowledgement:

"The project Music360 has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101094872.

The opinions expressed in this document reflect only the author's view and in no way reflect the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains."

Table of Contents

Management Summary

During the 4-year MUSIC360 project data will be collected from various sources, and from these data the project will generate new data of different types. Each type of data requires a different management practice.

This MUSIC360 data management plan is tailored to meet the specific needs of the project and its data. It guides how research data will be (i) collected, (ii) processed, (iii) analyzed, (iv) shared and (v) stored. It also provides guidelines for the management of different types of data. It adheres to the principles of the EU, to be "follows the principle "*as open as possible, as closed as necessary*" and focuses on encouraging sound data management as an essential part of research best practice. We follow in this report the EU guidelines in the **EU Data Management for Horizon 2020 [1].**

According to these EU guidelines, Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated. A DMP commonly includes information on:

- what data will be collected, processed and/or generated;
- which methodology & standards will be applied to collect, process and generate the data;
- the handling of the data during & after the end of the project;
- how data will be curated & preserved (including after the end of the project) [1].
- whether data will be shared/made open access

This data management plan is a living document, accessible for all partners. It will be regularly updated. In the present version, the first three chapters cover the general requirements for a data management plan. How this applies to the MUSIC360 project is described in chapters 4-6.

1. Introducing to Data Management and MUSIC360

1.1 Introduction to the project, the objectives, data sources and outputs

The MUSIC360 project aims to to uncover the real -- economic and social -- value of music. Background music, for example played in shops, bars, restaurants, is known to be of economic value. However, there is currently a lack of systematic and quantifiable obtained knowledge on the "real" use and value of music. This lack of knowledge may lead to an unfair distribution of the revenues of music among its creators and performers. This is the case in many countries across Europe. This may lead to uninformed decision-making at the European and national policy levels.

To uncover the real value of music, the MUSIC360 project will **collect, combine, process, interpret and make available data** about background music played at different venues, and its listeners. This is done in different European countries, in so-called Living Labs – real life venues where people are exposed to background music.

The MUSIC360 consortium comprises five CMOs, two universities, a fingerprinting company, a European lobby company for artists, and a company on ecosystem design. Jointly, their customers cover all relevant stakeholders in the EU music ecosystem.

The MUSIC360 project will develop a digital MUSIC360 Platform, still to be designed during the project according to requirements collected from the partners and the Living Labs. This platform will host data at a fine-grained level using digital **fingerprinting technology,** provided by one of the project partners: BMAT**.**

This data, generated during the project, will be combined with **data about composers, arrangers, performers and recordings, held by Collective Management Organizations**,

The aim is to make this data transparently available to stakeholders in the EU music ecosystem, including, creators, venues and policy makers. A dashboard in the Music360 platform will provide access to the platform in **a secure and personalized way** to various stakeholder groups: (i) creators and artists to understand the value of music and be able to identify new business opportunities; (ii) venue owners to assess the value of music for their revenue; (ii) policy makers to better understand the music ecosystem.

A theory of music value will be designed, based on real-world experiments with the CMOs and their customers. **A distributed architecture for the secure collection and sharing of data** will be developed together with the fingerprinting company.

From the outcome of the analysis of all data, a large-scale business model for the music industry will be designed, using conceptual modeling techniques such as e3-value[2], BPMN[3] and other conceptual modeling methods. Additionally, a conceptual model for the governance structure of the music industry will be designed.

In short, this research project deals with different types of data: pre-existing data from the music industry, newly collected data on background music from venues, data from interviews with music users, generated data as the result of various operations and analyses.

[2] The e3-value represents both a research methodology and a practical value modeling tool, see https://www.thevalueengineers.nl/download/e3value-tool/

[3] BPMN stands fof Business Process Model and Notation (BPMN); this is a standard for modeling of business work processes.

In Chapter 2 we will discuss the requirements for data management in general. In Chapter 3 we will discuss FAIR principles, because this is a protocol for opening up and sharing research data within a scientific community, however, this is not always possible for safety, privacy and ownership reasons.

2. Data Management requirements - in theory

2.1 Introduction to data management practices in general

A data management plan should describe for its research, what type of data is collected, processed and/or generated. Next, it must describe which methodology & standards will be applied when collecting, processing and generating data.

One important decision to take is whether to share data and make it Open Access, or keep it confidential. In the latter case security and encryption play a role. For Open Access, the question is whether FAIR principles are applicable. FAIR is further described in this chapter (see section 1.3).

Physical storage of data is also an important aspect, as it involves important decisions about e.g., provider, access method, backups etc. as discussed in the following sections. Also, it may incur costs.

For the future beyond the project period, it is important how data will be curated & preserved (including after the end of the project) [1].

2.2 Data collection in multi/trans-disciplinary research projects

Music360 is a multi- and trans-disciplinary project. Multidisciplinary in this project means that its research is based on different disciplines: Computer Science, Economics, and Social Sciences. Each of these disciplines collects and handles data in a different way, and it uses different methodologies and comes from different disciplinary traditions. This has influence on the way data will be handled and shared.

Trans-disciplinary research projects are done in collaboration between academic and non-academic staff. That is case in MUSIC360, in which universities collaborate with CMOs. The latter are data collecting organizations with their own data and own data management.

The CMOs offer various data resources to its members, e.g. concerning right owners on recordings and the works, the fees paid by right societies to right owners, and usage of music by users, e.g. in retail in the hospitality. Some data is open, while other data has serious confidentiality concerns. The data from the CMOs is not research data, but some parts of it will be used/processed by the project.

Therefore, this data management plan will have to be elaborated in such a way that it meets the different traditions and requirements.

2.3 About data classification

Data Management **data classification** is the process of categorizing and labeling data based on its sensitivity, value, and other attributes. It is a fundamental aspect of data security and -management what should be done, and this will also apply for the MUSIC360 project.

Public Data: Data that is intended for public consumption and doesn't require any special protection. This might include information that is already widely available or meant to be shared openly.

Internal Data: Data that is intended for internal use within the project but doesn't contain sensitive or confidential information. This data might include data that is not relevant for sharing.

Confidential Data: Data that is sensitive and requires protection. This category might encompass personally identifiable information (PII), financial data and proprietary business information.

Restricted or Highly Confidential Data: Data of the utmost sensitivity that requires the highest level of protection. This could include trade secrets, legal documents, medical records, and other highly confidential information.

2.4 Data sharing, data ownership and ethics

Legal & ethical considerations are applicable to all personal data and other types of sensitive data. Thus, it is important to consider data ownership and intellectual property of the data. For example, data from interviews is usually anonymized to protect the identity of the research subjects. However, there are type of research interviews, e.g., expert interviews, which are usual in requirements engineering and software development, in which the user is explicitly required to be mentioned with its full name. In such cases, the intellectual property of the interview may not be the ownership of the researcher. There are situations in which the data is co-created with research subjects. In that case the ownership may be shared. Therefore, ethical and legal considerations require that the researchers involve the stakeholders in the data management and make them aware of their roles and rights. This goes beyond what is often named as “informed consent”. A special case of data is the data that is subject to the European regulation called GDPR.

2.5 About General Data Protection Regulation (GDPR) and Data Protection Act

The General Data Protection Regulation (GDPR) is a comprehensive data protection and privacy regulation that was implemented by the European Union (EU) on May 25, 2018. The GDPR aims to provide individuals with greater control over their personal data and how it is collected, processed, and stored by organizations. It also establishes a standardized framework for data protection across the EU member states.

Privacy by Design: researchers are encouraged to integrate data protection into their systems and processes from the outset, and to ensure that only necessary data from individuals is collected and processed.

GDPR provides exemptions for personal data processing in relation to research activities. This means that the highest ethical standards and guidelines are rigorously applied, regardless of the country in which the research is carried out.

Participants will have the right to opt-out of any further processing without giving any motivation. In this case, personal data may not be erased but will only be used in an anonymized form as part of the dataset.

In accordance with accepted ethical standards, participants will not be named in any published materials unless they have given their explicit permission for this to happen. If participants would like to request a copy of the personal data, then they can contact the lead researcher.

Where practicable, they will provide participants with a copy of their data. Research data may be retained indefinitely in an anonymized form by researchers.

Key aspects and provisions of the GDPR, which are relevant for MUSIC360, include:

- Scope: The GDPR applies to all projects and organizations that process personal data, regardless of whether the processing occurs within the EU or not.
- Lawful Basis for Processing: The lawful bases include consent, contractual necessity, legal obligation, vital interests, public task, and legitimate interests.

- Consent: the lawful basis for processing personal data, implies that it is obtained in a clear and specific manner. Consent must be freely given, informed, and easily revocable.
- Data Subject Rights: The GDPR grants individuals several rights concerning their personal data, including the right to access, rectify, erase, restrict processing, data portability, object to processing, and not be subject to automated decision-making.
- Accountability and Transparency: It is required to be transparent about their data processing practices and to provide clear and concise privacy notices to data subjects.

2.6 Sensitive data, encryption, GDPR compliance

There is a category of data – sensitive data -- that must be treated with high confidentiality. How does the research project deal with data confidentiality? How does it protect sensitive data, and how is the anonymity of interviewees and other stakeholders ensured? To manage sensitive data, the following points must be addressed: Which data encryption will be used? How will data be monitored and audited to ensure that it is secure and complies to the rules and regulations?

2.7 Data processing; standardization and naming conventions

When the research data is being collected, we ask ourselves: What is the data quality, and how is this ensured? What are the methodologies for collection, and validation? How will the meta-data be managed – published or not? If data requires visualization: which visualization tools will be used and do these comply with all the above (security, privacy, IPR)?

Other points that need to be decided at the start of the data collection:

- Are standards used for file names and folder structure and organization to ensure interoperability within the project? Also, for variables, and vocabularies, aimed at interoperability.
- Listing of data files and repositories: how are they published or stored/documented?
- Are links to Open Data in other data sources made e.g. Linked Open Data, RDF, other standards.

2.8 Physical storage of the data, where, how big, how long?

It is important to decide at an early stage, how to manage the physical data. Will the collected data be hosted in the cloud? Does the hosting provider take care of the necessary facilities: backup, access via digitally secured access channels and secure redundant, distributed backup facilities? The following list will be further elaborated in the coming period of the project.

Repositories: which repositories will be used to make the data available (e.g., Github, AWS). What is the capacity needed in terms of storage? What are the costs involved? How can we make data public? Will the data be retained after the end of the project?

This is a checklist for the storage of the physical data:

- Capacity & costs of physical storage (amount needed)
- Redundancy (physical redundancy such as RAID)
- Physical security & accessibility protocols (data access for different categories of users)
- Backup & archiving procedures

- Data retention after the project period

3. Background to FAIR Data principles

The EU adheres to the principles of Open Science. This entails that all results of the EU-funded research projects are made public and published as Open resources. For data this is a wish, although data is not always Open for various reasons. Therefore, the idea is to manage data: "*as open as possible, as closed as necessary*" [1]. For Open data, the EU proposes, if possible, the FAIR principles:

According to FAIR principles, research data should in principle be made available to the scientific community, in such a way that it is **findable, accessible, interoperable, and reusable** [9, 10]. This implies that data is such that it can be processed by computational systems, enabling these to find, access, interoperate, and reuse data with none or minimal human intervention [11].

As pointed out, from the perspective of research sharing and disseminating, and for the purpose of replicability and scrutiny, FAIR is a good principle. However, in real world research, especially trans-disciplinary research, in which non-academics are involved in business, in which people’s privacy must be protected, and in which e.g. business partners have interests in protecting their business data, and in which GDPR law and regulation must be obliged, the principles of FAIR can often not be fully or partially met. Here's what each aspect of FAIR data entails (see [11])

1. **Findable**: Data should be easy to find and locate. This involves assigning globally unique and persistent identifiers to datasets, such as Digital Object Identifiers (DOIs) or Uniform Resource Identifiers (URIs). Additionally, metadata (descriptive information about the data) should be provided and made available in searchable repositories, making it easier for others to discover the data.
2. **Accessible**: Data should be accessible to both humans and machines. This involves providing clear and comprehensible access rights and permissions, ensuring that data are stored in a way that is secure and can be accessed without unnecessary barriers. Accessibility also includes providing appropriate documentation and contextual information about the data.
3. **Interoperable**: Data should be structured and formatted in a way that uses standardized data formats, ontologies, and vocabularies, as well as providing clear information about the data's structure and meaning. Interoperability ensures that data can be easily integrated and used across different research domains and disciplines.
4. **Reusable**: Data should be designed and documented in a way that allows it to be reused for various purposes. This includes clear licensing and usage terms, as well as providing comprehensive and accurate metadata that enable other researchers to understand and effectively use the data. Reusability also involves providing information on data provenance, methodologies, and associated workflows.

The FAIR principles have been set up in the context of Open Science and to facilitate collaborative research. By following FAIR principles, researchers can maximize the impact of their data and promote transparency and reproducibility in research. While the FAIR data principles offer numerous benefits there are some potential drawbacks with their implementation. The points below are considered for the MUSIC360project.

1. Privacy and Security: FAIR data principles emphasize open sharing and accessibility, which can raise concerns **about privacy and security**, particularly when dealing with sensitive or confidential data. This is the case for data collected in the MUSIC360 project, e.g. personal data, data from listeners, interviews, living labs. Not all data in the project is owned by the project. For example, the CMOs will continue managing their own data, as this is their core business. However, some data will be made available through the platform, and data will be used for modeling the value of music.
2. Disciplinary Differences: Different research fields have varying data practices, standards, and requirements. Applying FAIR principles uniformly across diverse disciplines can be challenging due to

differences in data types, formats, and research workflows. This is the case in the multi-disciplinary MUSIC360 project, in which the social sciences, economics and computer science are represented.

3. Long-Term Sustainability: Ensuring the long-term sustainability of FAIR data repositories and resources will be difficult. Maintaining data accessibility and interoperability over time requires ongoing investment in infrastructure and resources. Often, this cannot be sustained at the end of a project period.
4. Legal and Ethical Complexities: Different jurisdictions have varying legal and ethical considerations related to data sharing, intellectual property, and ownership. Navigating these complexities can be challenging, especially in international collaborations and the EU as in the MUSIC360 project.
5. Data Diversity: FAIR principles are designed to be applicable to a wide range of data types, but certain types of data, such as multimedia, complex models (such as e3-value) , or proprietary datasets, as will be used, collected and produced in MUSIC360, might pose unique challenges in terms of making them fully FAIR.

While these drawbacks exist, it's important to consider both the benefits and challenges of FAIR data principles. In the next chapters we will describe the data of the MUSIC360 project, and we will discuss to which level we can meet the FAIR standards in MUSIC360.

3.1 How to make FAIR Data operational

The following requirements are drawn from the VU Amsterdam DMP 2023.

Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision).
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used.
- Outline the approach towards search keyword.
- Outline the approach for clear versioning.
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how.

Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide the rationale for doing so.
- Specify how the data will be made available.
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited.
- Specify how access will be provided in case there are any restrictions.

Making data interoperable:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary and interoperability. If not, will you provide mapping to more commonly used ontologies?
- Increase data re-use (through clarifying licenses):
- Specify how the data will be licensed to permit the widest reuse possible.
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.
- Describe data quality assurance processes.
- Specify the length of time for which the data will remain re-usable.

Allocation of (financial) resources

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs.
- Clearly identify responsibilities for data management in your project.
- Describe costs and potential value of long -term preservation.

4. Managing data in the MUSIC360 project

As discussed in the previous chapter, a data management plan should meet certain requirements. In this chapter we will discuss this for the data of the MUSIC360 project. The project uses, collects, and processes data in different ways. First, we distinguish the following types of data in the framework of the MUSIC360 project:

4.1 Data categories in the MUSIC360 project

In the list below the different types of data are briefly described. For each of these categories, the data management that is applicable will be discussed.

Source data

Data stored and managed by the Collective Management Organizations (CMOs) in the project. This will be used to present an integrated view of music data. These data are owned by the CMOs and follow the rules as set by law to the management and storage. Since these are sources of data that will be used, measures will be taken when these data are being processed in the project. However, the primary data stay with the CMOs' responsibility, according to the existing regulations in the partner countries (Netherlands, Ireland, Finland, Portugal, Spain).

Fingerprint data

For the monetary value, the MUSIC360 project will use music fingerprinting to measure which background music is played in selected shops and restaurants and relate this to revenue data is. Storage and computation will be done locally at the partner's location, so that the data remains confidential. Fingerprinting is used in venues, currently focused mainly on recorded music. Specific hardware is installed in these venues. Currently this way CMO can get a more accurate picture of what music is being played outside of the broadcast-based data sources.

It is computationally not cost-effective to fingerprint all music played at every venue. So, it will always become a statistical approximation of reality, based on the gathered data. The choices made on when and in which venues fingerprinting takes place, is thus critical for the validity of the resulting approximation and a fair distribution or royalties.

Currently, some partner CMOs in the project are already using these data on venue playbacks in the royalty distribution process. There are currently around 1000 deployed devices installed, previous to the MUSIC360 project. An important consideration for CMOs in deciding to use fingerprinting or not, is its cost in relation to the distribution of royalties in the long tail, i.e. the music that is played less frequent. Costs per device are in the range of EUR 10-100 per month, depending on the number of devices deployed per client. Some CMOs offer discounts in licenses due by venues to incentivize venues in placing the devices.

BMAT also offers the fingerprinting technology through a smartphone app. This option adds flexibility and agility to the monitoring network, but the accuracy of the results is less due to limitations of the microphones in these devices, and they don't record 24/7. We are looking into crowdsourcing fingerprinting data, however there is, at the moment, an unsolved issue of trustworthiness of this data. Scaling up in this way also will cause some challenges in terms of computational cost.

Sensitive data from living lab experiments

Data collected in experiments with users in the so-called living-labs. These data are confidential and will be treated securely, using e.g. homomorphic encryption.

Software

There will be software developed in the project. This entails source code developed in the project, including documentation. Most software developed in MUSIC360 will be made available in Github under open-source licenses when no dependencies exist.

However, some software in MUSIC360 is developed as closed source (e.g., by partner TVE, as stated in the EU Grant and Consortium Agreements). Yet, there will be a community version of the software, delivered as Open Source, and an enterprise version that will be free for academic institutions but with a paid license for commercial business use.

Modeled data

Conceptual models and architecture of the value of music, of music data, will be produced as outcomes of the project research. These will be published in peer-reviewed scientific literature with open access license and the publications will be made public in the open science repositories of the universities UPV and VU, compliant to the EU Open Access guidelines [1]. Conceptual models of the distributed Platform architecture will be published in peer-reviewed scientific literature with open access license. Publications will be made available through Open Science repositories of the partner universities UPV and VU, compliant to the EU Open Access guidelines.

Personal data collected in the Living Labs

Personal data in various forms will be collected in the Living Labs experiments. This includes data collected from customers of venues, patients, and creators (performers, composers, text authors, artists, etc.). Data about customers will be fully anonymous, hence, are not identifiable in the analyses. Patient data on the effects of music in health care, in hospitals will only be collected by the researchers of UPV. Participation by patients is voluntary, for which a protocol of consent. Data storage will follow the ethical requirements for such data (e.g. limited access control and encryption). Processing data of music creators follows the practices and standards of their CMOs.

5. Data collected in the MUSIC360 project

5.1 Questions for each data cluster in the project

The following questions will be answered for the different data types in MUSIC

- the origin of the data.
- the purpose of the data collection/generation.
- the relation to the objectives of the project.
- the types and formats of data generated/collected.
- Specify if existing data will (possibly) be re-used (if any).
- State the expected size of the data (if known).
- Outline the data utility: to whom will it be useful.
- Possibly: data confidentiality and encryption, see below.

5.2 About data confidentiality and encryption of data in MUSIC360

There is one category of data – sensitive data -- that must be treated with utmost confidentiality. How does the MUSIC360 project deal with data confidentiality? How does it protect sensitive data, and how is anonymity of interviewees and other stakeholders ensured? To manage sensitive data, the following points must be addressed:

- Data encryption.
- Monitoring & auditing data.
- Compliance with GDPR and other relevant regulations.

In the WP 1, the researchers may want to do computations using data in the MUSIC360 Platform. This is data that is confidential to the stakeholder asking the query. Confidential data can be used in computations while delivering a non-confidential answer, e.g. the sum of the earnings per genre or per geographical region. The MUSIC360 Platform will be designed such that it is able to perform these computations without revealing the confidential data on which they are based. We will do this with (partial) homomorphic encryption techniques, which allow containerization of data in such a way that the data cannot be read directly, but can be used in e.g., calculations.

The VU has previously developed a partial efficient homomorphic encryption technique based on the Orlandi Protocol, see [6,7,8]. Currently, the work of the VU supports only arithmetic expressions. In the Living Labs we will further investigate and address this.

6. Data collected in the MUSIC360 project

6.1 Data collection for one of the CMO partners: BumaStemra, questions and answers:

- **What is the origin & purpose of the data (including metadata, conceptual models etc)?** *The origin of the data is the Author/Publisher? The provided data contains Personal data to identify Who is Who. It also contains Work information related to the created musical works, this to identify the musical works which are being used. The usage data is provided by music users or through technical solutions such as fingerprinting.*

- **How is the collection/generation/conceptualization done, and by whom?** *The members of each CMO are obligated to give the data to their own CMO. Thereafter each CMO is able to exchange this information among each other. The usage data is provided by music users or through technical solutions such as fingerprinting.*

- **What is the relationship of your data/models to the objectives of the MUSIC project? What are the types and formats of data generated/collected?** *All CMO's use industry standards, such as; ISWC (international standard works code, to identify the unique work); DDEX (data exchange format created for digital music exchange); CWR (common works registration, for exchanging work information among Author rights CMO's)'; IPI (Interested Party information, a unique identifier to establish who is who for Music Author CMO's)*

- **Which naming conventions do you use for your file names/variables?** *See previous*

- **Please specify if existing data will (possibly) be re-used (if any)?** *All Identity information and all Works information is being re-used.*

- **Please state the expected size of the data (if known)?** *Millions of Authors (worldwide) and tens of millions of works (worldwide).*

- **Please outline the data utility: to whom will it be useful?** *The Author and Publisher, and besides that in aggregated form also for music users.*

- **Is there data confidentiality/encryption required or is FAIR an option?** *All Identity information must be treated by the GDPR rules*

- **Data storage:** where, when, how, costs, how long? *In central databases (each CMO). Indefinitely for Identity and Works information. Just a few years for Music Usage information.*

6.2 Data collection for BMAT

- **What is the origin & purpose of the data (including metadata, conceptual models etc):** *The data is collected in the Living Labs, from the partners venues. The purpose is the identification of the music used and its enrichment with other data sources, mainly libraries with information such as genre or mood.*

Input: Audio streams from venues and associated metadata (date/time, venues address, other venues info like capacity, size, etc)

Output: Music usage reports with music metadata as a result from audio fingerprinting processing (date/time, duration, venue, song title, artist, label, ISRC) and other music metadata thanks to data enrichment processes if necessary (genre, work metadata, ...)

As developers of the dash-board we might receive other types of data provided by project partners as a result of the Living Labs data collection (e.g., interview results) or from their own databases (e.g., work documentation from CMOs), that will be processed and eventually added as is or aggregated in the dashboard.

- **How is the collection/generation/conceptualization done, and by whom?** *In most cases, BMAT will collect the data through the use of BMAT boxes, a device that records and sends the audio recording to BMAT premises, where the audio stream is processed with a fingerprinting solution that identifies its music content and outputs a report including music titles and related metadata from BMAT's music database. In specific cases, the list of songs will be provided by the Living Lab owner and enriched by BMAT with the necessary metadata.*

- **What is the relationship of your data/models to the objectives of the MUSIC project?** *The metadata attached to each music usage will allow the music used to be categorised and aggregated in a way that should allow the interpretation of the value of the music in each LL case when analysed together with data collected provided by other partners of the project.*

- **What are the types and formats of data generated/collected?** *See answer 1. Music usage reports. Data such as duration, channel, track, artist, label, isrc, iswc, bmatid, album, contributors, composer, music publisher, and additional metadata such as genre, mood, vocal or instrumental, language... Music usage reports are text files that can be generate in tabular format such as CSV or .xlsx***.** *Music usage information can also be accessed via API*

- **Which naming conventions do you use for your file names/variables?** *This will be based on the format of the names given to audiostreams, but still has to be defined for the project together with the LL partners.*

- **Please specify if existing data will (possibly) be re-used (if any)?** *No, except BMAT proprietary metadata used for enrichment purposes.*

- **Please state the expected size of the data (if known) ?** *To be defined.*

- **Please outline the data utility: to whom will it be useful ?** *The data will be useful to the Living Labs owners, to the university (to evaluate the value of music) and to the venues involved in the Living Labs*

- **Is there data confidentiality/encryption required or is FAIR an option?** *It will depend on the Living Lab partner, but probably confidential, although we can probably produce an open dataset from the Spanish Living Lab, from the Valencia hospital or "Moros y Cristianos" experiments. It is possible that we can share an open dataset with aggregated data*

- **Data storage: where, when, how, costs, how long?** *BMAT can store the audio streams for the course of the project and maybe even later if necessary. BMAT can store for as long as needed the* *results of the fingerprint processing and enrichment, including the original audio fingerprints from the audio streams.The cost depends on the volume of the dataset, which is not yet known.*

References and further reading

[1] EU Data Management for Horizon 2020

https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[2] Wyatt, Sally, et al. "Illness online: Self-reported data and questions of trust in medical and social research." *Theory, Culture & Society* 30.4 (2013): 131-150.

[3] Van Reisen, M., Stokmans, M., Basajja, M., Ong'ayo, A. O., Kirkpatrick, C., & Mons, B. (2020). Towards the tipping point for FAIR implementation. *Data Intelligence*, *2*(1-2), 264-275.

[4] Michener, W. K. (2015). Ten simple rules for creating a good data management plan. *PLoS computational biology*, *11*(10), e1004525.

[5] Yu, L., & Yu, L. (2011). Linked open data. *A Developer's Guide to the Semantic Web*, 409-466.

[6] T. Jakobsen, M. Makkes and J. Dam Nielsen, "Efficient Implementation of the Orlandi Protocol Extended

Version," 2010. [Online]. Available: https://eprint.iacr.org/2010/224.pdf.

[7] T. Jakobsen, M. Makkes and J. Dam Nielsen, "Efficient implementation of the Orlandi protocol," in Applied

Cryptography and Network Security, Berlin, 2010.

[8] M. Makkes, A. Uta, R. Bharath Das, V. Bozdog and H. Bal, "P^2-SWAN: Real-Time Privacy Preserving

Computation for IoT Ecosystems," in 1st International Conference on Fog and Edge Computing (ICFEC), 2017.

[9] Mark D. Wilkinson; Michel Dumontier*; IJsbrand Jan Aalbersberg; et al. (15 March 2016).* "The FAIR Guiding Principles for scientific data management and stewardship". Scientific Data*. 3 (1): 160018.* doi*:*10.1038/SDATA.2016.18. ISSN 2052-4463. PMC 4792175. PMID 26978244. Wikidata Q27942822*.*

[10] Annika Jacobsen; Ricardo de Miranda Azevedo; Nick Juty; et al. (31 January 2020). "FAIR Principles: Interpretations and Implementation Considerations". Data Intelligence: *10–29.* doi*:*10.1162/DINT_R_00024. ISSN 2641-435X. Wikidata Q76394974*.*

[11] "FAIR Principles". *GO FAIR*. Retrieved 2020-02-16. Material was copied from this source, which is available under a Creative Commons Attribution 4.0 International License.

Annex 1

In this annex we briefly show the VU data management plan[4]. According to the VU Data Management Protocol, a DMP consists of various elements as discussed in the previous sections. In the MUSIC360 project we use this for further elaboration of this DMP. The discussion about whether to comply to FAIR, see the EU DMP Guidelines [1], leaves room for deviations, and amendments. Therefore, we will use the VU checklist not as a protocol, but as a recommendation. The text following is copied from an internal document of VU by Kees Verstoep (c.verstoep@vu.nl) and Shuai Wang (shuai.wang@vu.nl) September 2023, titled: "Selecting the right storage platform for VU Computer Science research projects "

By Kees Verstoep (c.verstoep@vu.nl) and Shuai Wang (shuai.wang@vu.nl) September 2023. In the online VU tool Data Storage Finder no less than 10 storage options are listed. They are classified as appropriate or not according to the following criteria:

Data classification (sensitivity)	low / medium / high / very high
Data sharing	not needed / with VU colleagues / with anyone
Data volume	below 500 GB / above 500 GB
Possible features	- Basic storage - Compute/HPC - Fine-grained access rights - Collaboration tools - Archiving/Publishing

Using many different storage options can lead to fragmentation and operational overheads when working with multiple data sources. Therefore, in our department we want to limit the commonly used storage options to a preferred shortlist. This document provides this shortlist.

Apart from the options mentioned, personal laptops and storage tied to data processing resources like HPC clusters are of course also important data management resources for research. However, these should only be used to work on (copies of) the actual datasets that are stored on and shared via the selected primary storage platform.

Version management of primary and derived research data is an important topic in itself which needs to be addressed in a project's data management plan, and which should be made explicit in operational guidelines which can also include suitable synchronization mechanisms.

[4] See also the VU website for DMP https://dmponline.vu.nl/plans