Polifonia: a digital harmoniser for musical heritage knowledge, H2020

D6.4 Policy Recommendations, version 1

Deliverable information	
WP	WP6
Document dissemination level	PU Public
Deliverable type	R Document, report
Lead beneficiary	NISV
Contributors	UNIBO, KCL, KNAW
Document status	Final
Document version	1.0
Date	20/12/2021
Authors	Jesse de Vos (NISV, task leader) Valentina Presutti (UNIBO) Albert Merona Penuela (KCL) Marilena Daquino (UNIBO) James McDermott (NUI GALWAY) Marco Gurrieri (CNRS) Enrico Daga (OU) Andrea Scharnhorst (KNAW) Nicole Emmenegger (NISV)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004746

INTENTIONALLY BLANK PAGE

Project information

Project Start Tate: 1st of January 2021

Project Duration: 40 months

Project website: http://www.polifonia-project.eu

Project contacts

Project Coordinator

Valentina Presutti

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
Department of Language, Literature and Modern Cultures (LILEC)

E-mail: valentina.presutti@unibo.it

Project Manager

Marta Clementi

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
Research division

E-mail: marta.clementi3@unibo.it

POLIFONIA consortium

No.	Short name	Institution name	Country
1	UNIBO	ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA	Italy
2	OU	THE OPEN UNIVERSITY	United Kingdom
3	KCL	KING'S COLLEGE LONDON	United Kingdom
4	NUI GALWAY	NATIONAL UNIVERSITY OF IRELAND GALWAY	Ireland
5	MiC	MINISTERO DELLA CULTURA	Italy
6	CNRS	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS	France
	SORBONNE	SORBONNE UNIVERSITE (LinkedTP)	France
7	CNAM	CONSERVATOIRE NATIONAL DES ARTS ET METIERS	France
8	NISV	STICHTING NEDERLANDS INSTITUUT VOORBEELD EN GELUID	Netherlands
9	KNAW	KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN	Netherlands
10	DP	DIGITAL PATHS SRL	Italy

Project Summary

European musical heritage is a dynamic historical flow of experiences, leaving heterogeneous traces that are difficult to capture, connect, access, interpret, and valorise. Computing technologies have the potential to shed a light on this wealth of resources by extracting, materialising and linking new knowledge from heterogeneous sources, hence revealing facts and experiences from hidden voices of the past. Polifonia makes this happen by building novel ways of inspecting, representing, and interacting with digital content. Memory institutions, scholars, and citizens will be able to navigate, explore, and discover multiple perspectives and stories about European Musical Heritage.

Polifonia focuses on European Musical Heritage, intended as musical contents and artefacts - or music objects - (tunes, scores, melodies, notations, etc.) along with relevant knowledge about them such as: their links to tangible objects (theatres, conservatoires, churches, etc.), their cultural and historical contexts, opinions and stories told by people having diverse social and artistic roles (scholars, writers, students, intellectuals, musicians, politicians, journalists, etc), and facts expressed in different styles and disciplines (memoire, reportage, news, biographies, reviews), different languages (English, Italian, French, Spanish, and German), and across centuries.

The overall goal of the project is to realise an ecosystem of computational methods and tools supporting discovery, extraction, encoding, interlinking, classification, exploration of, and access to, musical heritage knowledge on the Web. An equally important objective is to demonstrate that these tools improve the state of the art of Social Science and Humanities (SSH) methodologies. Hence their development is guided by, and continuously intertwined with, experiments and validations performed in real-world settings, identified by musical heritage stakeholders (both belonging to the Consortium and external supporters) such as cultural institutes and collection owners, historians of music, anthropologists and ethnomusicologists, linguists, etc.

Executive summary

The policy recommendations in this deliverable are based on the work and experiences within the Polifonia project in the first twelve months. This report identifies areas of policy that affect the efficiency and quality of work that can be done by initiatives that engage with and promote European musical heritage. Its purpose is to inform and engage with policy makers at various levels to address issues relevant to Polifonia's work. These recommendations are structured in accordance with the FAIR principles, which ensure that data is Findable, Accessible, Interoperable and Reusable.

Document History

Version	Release date	Summary of changes	Author(s) - Institution
V0.1	11/11/2021	First draft for review	NISV (Jesse de Vos)
V0.2	07/12/2021	Edited based on review Andrea Scharnhorst and Nicole Emenegger	NISV (Jesse de Vos)
V0.3	15/12/2021	Edited based on final review Andrea Scharnhorst and Nicole Emenegger	NISV (Jesse de Vos)
V1.0	20/12/2021	Final version submitted to EU	UNIBO

Table of contents

Project information 3
Project contacts 3
POLIFONIA consortium 3
Project Summary 4
Executive summary 5
Document History 6
Table of contents 7
1. Introduction 8
1.1 Purpose of deliverable 8
1.2 Work done in Polifonia thus far 8
2. Policy recommendations 10
2.1 Findability 11
2.2 Accessibility 12
2.3 Interoperability 13
2.4 Reusability 15
3. Conclusions 18
4. References 19

1. Introduction

1.1 Purpose of deliverable

If one thing has become apparent during the pandemic, it's the notion that digital engagement in society is still developing. People's experiences have been more confined to digital, online spaces and their interactions with cultural heritage has followed suit. As researchers and curators, we believe that our cultural heritage is both a reflection of values in society and a means to critically reflect on those values. It is also a source of beauty and entertainment and we have a responsibility to engage with people wherever they are. It has become paramount to improve the online presence of (in this case) musical cultural heritage: to organise our collections, datasets and research results in such a way that they can be retrieved, interpreted and reused.

The policy recommendations in this deliverable are based on the work and experiences within the first twelve months of the Polifonia project. The report identifies areas of policy that affect the efficiency and quality of work that can be done by initiatives that engage with and promote European musical heritage. Logically, this deliverable is placed in work package six, which concerns itself with dissemination and engagement aspects of Polifonia. With this deliverable we wish to inform and engage with policy makers at various levels to address issues relevant to Polifonia's work. The project consortium is especially concerned with connecting and enriching objects of musical heritage and this document will therefore address issues that relate to the activities of publishing data about musical heritage.

The recommendations in this report are mostly aimed at policy makers that work for organisations managing and publishing heritage data, and the organisations that are responsible for regulation in these areas. This includes, but isn't limited to, policy makers at cultural institutions, people involved with research data management in universities, but also people in the political sphere that can influence the agenda of ministries of culture, and legal experts that are involved with broadening possibilities for research of and engagement with cultural heritage.

This version of the policy brief is a first exploration of policy and issues that affect the efficient distribution and linking of musical heritage. Towards the end of the project a final policy brief will provide a more comprehensive overview that is based on the experience of the Polifonia project in its entirety.

1.2 Work done in Polifonia thus far

The work in Polifonia is mostly organised around ten pilots, that drive the development of the ecosystem through continuous validation of its technologies. These pilots address various topics related to improving access to musical cultural heritage. An overview of these pilots can be found in table 1 below.

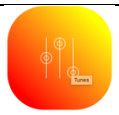
Preserving musical heritage through knowledge graphs			**https://polifonia-project.eu**
1	BELLS	Preservation of Historical Bell Heritage: dependencies between tangible and intangible	
2	ORGANS	A Knowledge Graph on History of Pipe Organs	
Managing musical heritage collections through knowledge graphs			
3	FACETS	Exploration of music scores collections through statistical features	
4	INTERLINK	Interlinking of collections in digital music libraries and audiovisual archives	
Studying musical heritage through (interlinked) knowledge graphs			
5	CHILD	Exploration of musical heritage for scholarly enquiry: a case study on Music and Childhood.	
6	MUSICBO	Knowledge graph of Bologna Musical Heritage.	
7	TUNES	Tunes analysis and classification	
8	TONALITIES	Modal and tonal classification of Western notated music from the Renaissance to the 20th century	
Interacting with musical heritage knowledge graphs			
9	ACCESS	Making musical performances accessible to people who are Deaf or hearing impaired	
10	MEETUPS	Musical Meetups: the European musicianship flow	

Table 1: Polifonia's *10 Pilots:* Label and description, Icon and Grouping[1]

[1] For more detailed descriptions see Polifonia Deliverable 7.1 Data Management Plan. June 2021.

Each of the pilots generates data which will be connected in the Polifonia Web Portal as described in the deliverable D1.3.[2] In the web portal already existing and new datasets developed in the pilots are published, and where possible, connections between them are made. The number of datasets available in the web portal will grow over time, not just by adding datasets from the Polifonia project, but also from initiatives outside of the project. In this central hub the datasets can be queried and relations between them found, forming a distributed knowledge graph. In addition, a dataset catalogue is being developed to improve discoverability of the existing datasets themselves.[3]

This distributed ecosystem created by Polifonia allows all kinds of new interactions with musical heritage. To exemplify these interactions Polifonia presented a demo at the SONAR festival in October of 2021. In it, users were presented with an audio file of a current pop song. As the song is playing, a feed provides all kinds of automatically generated enrichments, based on analysis of the lyrics of the song, metadata about the song (where it was recorded, who is playing, when it was performed) and similarities to other music (e.g. melody lines occurring in classical music, harmonic patterns occurring in other songs of the same band, etc.). This demo was of course tailored towards a general audience, but one can imagine how these same links can be used by researchers and curators to establish all kinds of new and interesting relations between musical heritage over time and across borders. By doing so the combined European cultural history of musical development can be explored.

2. Policy recommendations

The policy recommendations below are based on the experiences in the Polifonia project thus far. They highlight a number of issues, namely around findability, accessibility, interoperability and reusability of both relevant datasets and developed tools and explore ways to address these issues at a policy level. The issues mentioned largely follow the FAIR principles for scientific data management and stewardship.[4] The Polifonia project clearly displays the importance of these principles, and shows how these principles should not only guide the publication of scientific data, but also data from the cultural heritage sector. However, in this policy brief we are not looking to replicate the work done by current projects and networks such as GO FAIR[5] or FAIRsFAIR[6], in further detailing implementation issues of the FAIR principles. Instead, we will zoom in on a number of concrete issues we face in the Polifonia endeavour. This way, this report also complements the discourse around how to best implement FAIR with observations of concrete problems during implementation. The report is organised around identified *issues* and *recommendations*. In the recommendation subsection one example we have come across within the Polifonia-endeavour.

[2] Polifonia Deliverable D1.3 Pilots development – collaborative methodology and tools, June 2021

[3] This dataset catalogue is an improvement of the already existing Musow catalogue: https://musow.kmi.open.ac.uk/

[4] https://www.go-fair.org/fair-principles/

[5] https://www.go-fair.org/fair-principles/

[6] https://www.fairsfair.eu

2.1 Findability

Issues

In setting up the described ecosystem which supports the discovery of non-trivial relations between objects of musical heritage across borders of countries, institutions and disciplines, the availability of datasets is crucial. Polifonia's work in developing a dataset catalogue for such datasets brought to light a number of issues. Metadata about datasets isn't published centrally or consistently, which makes it difficult to find and therefore requires a lot of manual labour to gain insight into the available relevant datasets out there on the web. Once datasets have been found, it is hard to know what these datasets are about, and what their relationship is to musical heritage. Most dataset annotations and metadata are generic and do not cover specific subdomains of musical heritage. Also often unclear are the source and provenance of a dataset. This became in particular visible in the survey conducted by work package one[7], and the 'data stories' interviews conducted as part of the Research Data Management Plan.[8]

This unclarity about the source and provenance isn't only true for datasets that contain metadata and content. The same goes for ontologies and vocabularies that have been developed to represent musical heritage information.[9] Simply resorting to web search, dataset search and querying a public data repository such as Zenodo do not render satisfactory results. They provide little insight into what domains and questions ontologies cover, which datasets have been annotated with them and which knowledge graphs employ them.

Recommendations

- Put in the effort of publishing relevant subsets to users

Publishers of data should be made aware of their responsibility in publishing their datasets in such a way that these can easily be retrieved. Many cultural heritage institutions might publish their catalogue metadata, covering the entirety of a collection, but not bother with describing subsets of their catalogue related to, in our case, music. This means that the user of the data either has to be aware of relevant subsets and do the work of extracting the relevant subsets themselves from the full catalogue data, or that users might not be able to find this data at all. Providing the subset at the source means that expert curators familiar with the dataset can determine which resources are related to musical history and relevant tags can be attributed to a subset that are meaningful for expert users from the musical research community. Alternatively, example queries can be provided that return relevant subsets to users.

- Use public data repositories

The same goes for research data and the ontologies developed by researchers to analyse the data. After having finished a journal publication or research project, the data and ontologies should be published in public data repositories, such as Zenodo or Kaggle, or a specific dataset registry such

[7] Polifonia Deliverable 1.1
[8] Polifonia Deliverable 7.1 Research Data Management Plan
[9] See for those ontologies developed by the Polifonia project Polifonia Deliverable D2.1 Ontologies and knowledge graphs of music objects, patterns, and software package – 1st version, December 2021

as the musoW registry[10] which is developed within the Polifonia project. Datasets should be attributed with sufficient, accurate and fine-grained metadata for future users to be able to retrieve and reuse it. Protocols should be established that both heritage institutions and researchers adhere to in publishing their datasets and the use of standard vocabularies should be encouraged so that standardized terms can be used to query.

Example 1: The musoW registry

musoW is a catalogue of musical resources available on the web realized with the idea to support teachers in music education, creative industries, historians, and musicologists in finding what they need.

Usually, scholars and creatives need to combine diverse resources (music scores, audiovisual materials, data) from digital music libraries and audiovisual archives. They need to identify valuable sources of information, find similarities or cross-references. Tech-savvy people want to access music data programmatically to develop their innovative projects. The research of good music sources is mostly done manually, reviewing tons of websites.

In musoW, Polifonia supports users in searching across hundreds of online resources, including digital libraries, archives, datasets, and related technologies.

2.2 Accessibility

Issues

Datasets, once found, exist in all kinds of distributions on the web. In their most basic form they exist as a data dump in (hopefully) a non-proprietary format. In more ideal situations it is published via an API or harvesting protocol such as OAI-PMH, or it's available on a SPARQL endpoint. However, the plethora of types of endpoints makes it harder to synchronize between these datasets. Due to copyright a lot of musical heritage has limited (public) availability. This topic is further addressed in paragraph 2.5 Reusability below. But this is a reality in which researchers in this field function. It seems though that access for researchers to relevant datasets is often more restricted than necessary. There are all kinds of exceptions and agreements with copyright holders that allow datasets to be published in environments that require some sort of authentication. These datasets are hidden from plain sight and therefore harder to find.

Recommendations

- Use standard distribution methods

Make sure that data is published in such a way that no proprietary tools or communication methods are needed, and where licenses allow it, are publicly accessible. When creating endpoints on datasets, try to conform to what is commonly used within a certain field to make these more accessible. Provide example queries that give people unfamiliar with the dataset ideas on how to search within the data and which kinds of questions can be answered based on the data.

- If authentication is required, still focus on accessibility

[10] https://projects.dharc.unibo.it/musow/

If the material must be published under condition of user authentication, try to connect to existing platforms that facilitate such access so that the material is available to as many existing users as possible right away. Make sure the availability of the dataset is publicly known, so that people that qualify for an account can find their way to them. The dataset metadata should be registered in relevant dataset registries (e.g. Dutch Heritage datasets in the NDE Dataset registry[11] or European musical heritage datasets in musoW[12]), and the metadata describing the dataset should contain information on the degree to which and to whom it is accessible. Accounts to access these datasets should be easy to acquire for a broad range of researchers, not just from academics, but also for curators.

Example 2: Providing example queries on Wikidata

It may not always be possible, or desirable, to publish specific datasets for specific uses. Users of large datasets however may struggle to understand the data and the type of questions they can answer using that data. Providing a SPARQL endpoint with a user-friendly interface and example queries can go a long way in inspiring and informing users about the possibilities. The SPARQL endpoint of Wikidata[13] for instance, provides both example queries and a query generator. The resulting queries can easily be adopted to cater to the users' needs.

2.3 Interoperability

Issues

In order to find meaningful connections between datasets there have to be sufficient ways in which to make these connections. Objects of musical heritage are varied, ranging from sheet music, audio recordings, performance registrations and musical instruments to contextual archival records such as letters and other documents. Added to that there are datasets that are extracted out of audio files and sheet music based on content analysis, such as melodic, harmonic and rhythmic patterns, either manually or by using algorithms. All this information often exists in separate datasets without exposing the relationships between them.

In order to perform analytical tasks across these datasets a lot of work goes into preparing the data, e.g. conforming it to a single dataformat or annotating it with standard vocabulary. This work is usually done once and then not published sustainably for future reuse.

When new datasets are created based on existing datasets, the relationship between the two becomes unclear. Any changes in the original dataset are not transferred to any derivative or enriched datasets. This causes a messy landscape of partially overlapping datasets that can no longer be traced back to their source.

Next to datasets and ontologies, the ecosystem Polifonia is developing consists of a number of tools and technologies that, once integrated, will play an important role in answering the needs of relevant communities and stakeholders. Since it is a collaborative initiative, these technologies originate from a diverse range of partners. Coordinating the technical developments and resulting output components is challenging. Typically, research outputs are scattered in different repositories, using different technologies (e.g. programming languages and frameworks). Integrating the systems under

[11] https://datasetregister.netwerkdigitaalerfgoed.nl/
[12] https://musow.kmi.open.ac.uk/
[13] https://query.wikidata.org/

a single framework or technology constitutes a significant overhead for the project and usually results in assets that are hard to maintain and reuse after the end of the project.

Recommendations

- Dynamic dataset metadata and documentation

To address these issues and subsequently increase interoperability, dataset descriptions should evolve over time, including information about how they relate to other datasets. Information that is too detailed to be captured in descriptive metadata fields should be documented and datasets. This documentation should include information about which datasets it has been linked to in the past and which methods were used in order to do so. It should also contain information about previous experiments that have been performed with the dataset as well as any publications that made use of these datasets.

- Persistent identifiers

A more radical and comprehensive way of preventing the messy landscape of datasets described above, but harder to implement, would be to adhere more radically to Linked Data standards. A distributed network in which datasets and data points within the data can be referenced using persistent identifiers means that once connections are established, these can be sustainably managed, reused and expanded upon. Therefore, the use of persistent identifiers should be encouraged in libraries, archives and museums, but also in the research community. Persistent identifiers should be used at various levels, they can refer to collections, datasets, records, statements about records, descriptive terms, but also authors (e.g. ORCID's).[14]

- Shared ontologies and metadata formats

To make the process of linking the wide range of available datasets easier, the use of standard ontologies is mandatory. Granted, for specialised analysis specific ontologies, or extensions of ontologies can be required. This is work that the Polifonia project contributes to[15]. But the most basic information about musical heritage objects, that answers the questions of who, where and when, can be modelled in generic, widely used ontologies such as schema.org, which will greatly improve the interoperability.

- Shared vocabularies

Data can be described, tagged and annotated with terms from existing standard vocabularies. If all datasets were to use the same vocabularies to describe assets, this would of course be ideal. Often the choice for a particular standard vocabulary however is based on the vocabulary most commonly used in a specific sector or around a specific topic at a certain time. And once a choice for a specific vocabulary is made to describe items in a certain collection, it is unlikely that this will change. Creating alignments between vocabularies, where terms from one vocabulary are referenced in the terms of another vocabulary, makes seeing relations between datasets using those vocabularies much easier.

- Modular project outputs

[14] For more detailed information about the use of PIDs we refer to the FREYA project funded under the Horizon 2020 Research and Innovation Programme (https://project-freya.eu/)
[15] See for instance Polifonia Deliverable D2.1: Ontology-based knowledge graphs for music objects

Within Polifonia we follow a Knowledge Engineering approach to orchestrate the plethora of components developed or brought in by technical partners. We do this by means of an Ecosystem aimed at specifying the role and function of each asset (software, data, documentation, tutorials, requirements, …) in the context of the pilot use cases. The goal is to govern the diversity of assets without enforcing a single framework or technology, therefore reducing the effort of system integration. Instead, pilots will be developed by composing the various assets in pipelines, leveraging existing standards for interoperability, building on the tradition of Web technology specifications, Linked Data standards, Ontologies, Open Source software development, and good practices in the development of distributed systems.[16] Policy makers should encourage research projects to modularize project outputs, specify how they are used together to satisfy the project needs and document how stakeholders involved in developing end-user applications can reuse them beyond the scope of the project. The Polifonia Ecosystem can constitute a proof of concept of a novel approach to orchestrating project outputs and maximising their quality in terms of software and data engineering, and reduce the risks of the assets degrading after the end of the project.

Example 3: Shared vocabularies in a 'Network of Terms'

The reality of using vocabularies is that these are often tailored towards use for a specific collection and are hidden from sight for other potential users of the same terminology. In the Netherlands cultural heritage organisations worked together to build a 'network of terms'[17] in which various vocabularies can be queried at the same time. This provides a much better picture of coverage of openly available terminology sources. In addition, the network has created technology[18] to make alignments between SKOS-vocabularies, which makes it possible to cross-reference linked terms across multiple collections.

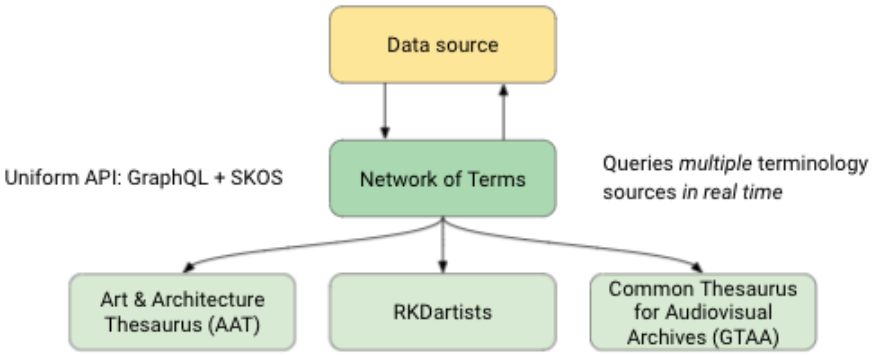


Image: an overview of the Network of Terms created by the Dutch Network for Digital Heritage.

2.4 Reusability

Issues

One of the most pressing matters within the field of research into musical heritage is the issue of copyright. Copyright can present severe limitations on what information a researcher or curator is able to find online and if they do find it, what they are able to do with this data in their analysis.

[16] More on this methodology can be found in Polifonia Deliverabl 1.3: Pilots development - collaborative methodology and tools.

[17] https://termennetwerk.netwerkdigitaalerfgoed.nl/

[18] https://cultuurlink.beeldengeluid.nl/

Especially when analysing collection at scale, using digital methods, it simply becomes too expensive to buy licenses for large amounts of copyright-protected material from publishers or copyright organisations.
One of the first issues that we run into is the lack of license information provided with music-related datasets. This lack of clarity almost necessarily means that by default copyright protection is assumed and therefore the data can't be used. Digital music files (sheetmusic, audio files, etc.) are themselves subject to copyright and related rights like the executive rights of performing musicians. The purpose of these laws is to make sure that creators can protect their creative works against theft and receive proper remuneration. As a consequence, research activities may be limited because of it as well. Music recordings, lyrics and sheet music may not be freely available on the web. And if the music is available, analysis of such files might be subject to copyright law. There is a grey area in the degree to which content analysis of music, and forthcoming metadata is a copy or derivative of the original work, and therefore subject to copyright law.

Recommendations

- Explore industry-research collaborations

Music publishers should be encouraged to support researchers by allowing access to their catalogues. Schemes of Return on Investment (ROI) should be analysed and identified to establish mutual benefits for industry, research and heritage. Polifonia for instance is committed to publishing the results of our analysis as open data (following the FAIR principles here described). We always link to the original source (e.g. the owner of recorded musical work) and share the results with the data owner. This can help them to exploit these analyses, for instance in promoting their music, making better recommendations and understanding their own ties to music culture and history. Owners of data can thus become early adopters of innovative technology to reach relevant audiences. It is important to note though that collaborations between research and industry do no solve all issues because not every type of research renders outcomes useful to the industry.

- Provide clear copyright information in dataset descriptions

Copyright information should be clearly stated in dataset descriptions. This information concerns both copyright on the dataset itself, as well as the assets it contains. In describing copyright status it is most useful to use existing standards and machine-readable statements (such as Creative Commons and Rightsstatements.org). In attributing licenses, owners of data should be aware of the restrictions their chosen license poses for users of the data. In general licenses will determine the degree to which users can share and adapt, but also whether they have to attribute (mention the owner of the dataset) and share their own adaptations of the dataset under the same license (share alike). It is easily underestimated how restrictive a certain license can be. For instance, having to attribute the owner of the datasets makes it very cumbersome for third parties to reuse the data as Linked Data. This is why for example Europeana has chosen the Creative Commons CC0 license. Research projects that are publicly funded should embrace open licensing as a default mode of work in order for the results to be as widely available to the public as possible, and in addition to allow researchers to build on the work that has already been done.

- Invest resources towards investigation of copyright status of existing collections

Heritage institutions and other collection holders should invest time and resources into exploring the degree to which their collections are available in the public domain. The knowledge in the Polifonia

knowledge graph has the potential to contribute to this exploration. If works can be linked to sources containing the death date of authors (composers), material can confidently be published in the public domain if the copyright has expired.
In addition, this research can provide a clearer picture of who rights holders are and agreements can be made with them for access, even if just for educational and research purposes. Institutions should be made aware of exceptions under copyright law that they can benefit from in providing access to their collections.
Finally, it is also worthwhile exploring the possibilities for extracting information out of music-related collections and the degree to which this derivative information is exempt from copyright law.

- Define template agreements

Define template agreements that can be used by European research project consortia to approach publishers and other rights holders. These agreements would address intellectual property issues, publishing/sharing limitations, allowed use, etc. Each European project can adjust these templates according to their needs, but in this way the main effort of establishing the terms is done once and can then be replicated.

Example 4: Europeana Licensing Framework and Guidelines

Europeana shows that large scale open access to heritage objects and their metadata is possible and has considerable experience in dealing with the many issues that come up in convincing stakeholders to make their data available under open licenses.

In the Europeana Licensing Framework a clear distinction is being made between metadata and the contents that are being described by the metadata. The metadata itself must be available to Europeana under the CC0 license, however, data providers can decide themselves how rich the metadata they provide to Europeana should be. This allows data providers to keep certain types of metadata to themselves with the exception of some mandatory metadata fields. Where the objects described are in the Public Domain, drawing on the Europeana Public Domain Charter[19] the Europeana Licensing Framework requires that data providers label digital objects that are in the public domain as such by applying the Public Domain Mark to them. To promote responsible re-use of public domain works that includes attribution for institutions that have invested in preserving, digitizing and making them available, Europeana has developed Usage Guidelines[20] for Public Domain works.

[19] https://www.europeana.eu/en/rights/public-domain-charter
[20] https://www.europeana.eu/en/rights/usage-guidelines-for-metadata

3. Conclusions

The above provides a generic overview of the most pressing issues that negatively impact the work Polifonia is trying to do. Polifonia will work to address these issues within the project, by actively engaging with communities of dataset publishers, to get as many relevant datasets within the Polifonia Knowledge Graph as possible. However, many of the issues listed above are beyond the influence and scope of our project. These issues are only properly addressed by a continued, long-term focus by funding bodies at various levels, policy makers nationally and at a European level and higher management of institutions.

In a future version of this policy brief (due in December 2023) more detailed recommendations will be made based on the experiences with the Polifonia project: which metadata standards, ontologies and vocabularies to use, how to leverage copyright exemptions for research purpose and how to increase interoperability. When further advanced in the project we will be able to provide more examples of where the issues mentioned are most pressing. The next version will also tailor the recommendations to more specific policy makers and by doing so make them actionable in the hopes of strengthening the European musical heritage information infrastructure overall.

4. References

- Bottini, Thomas et al. Polifonia Deliverable 1.1, Roadmap and pilot requirements. 2021

- Admiraal, Femmy. Andrea Scharnhorst et al. Polifonia Deliverable 7.1, Data Management Plan. 2021

- Daga, Enrico et al. Polifonia Deliverable D1.3 Pilots development – collaborative methodology and tools, 2021

- Merono Penuela, Albert. Polifonia Deliverable D2.1: Ontology-based knowledge graphs for music objects. 2021