



## D1.2 Roadmap and pilot requirements 2nd version

Christophe Guillotel-Nothmann, Jacopo de Berardinis, Thomas Bottini, Philippe Cathé, Enrico Daga, Marilena Daquino, Achille Davy-Rigaux, Marco Gurrieri, Peter Van Kranenburg, Eleonora Marzi, et al.

### ► To cite this version:

Christophe Guillotel-Nothmann, Jacopo de Berardinis, Thomas Bottini, Philippe Cathé, Enrico Daga, et al.. D1.2 Roadmap and pilot requirements 2nd version. Institut de recherche en musicologie (IReMus). 2022. halshs-04079523

**HAL Id: halshs-04079523**

**<https://shs.hal.science/halshs-04079523>**

Submitted on 28 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



*Polifonia: a digital harmoniser for musical heritage knowledge,  
H2020*

**D1.2 Roadmap and pilot requirements 2<sup>nd</sup> version**

Deliverable information	
WP	WP1 Pilots and Web portal
Document dissemination level	<ul style="list-style-type: none"><li>○ PU Public</li></ul>
Deliverable type	<ul style="list-style-type: none"><li>○ R Document, report</li></ul>
Lead beneficiary	CNRS
Contributors	UNIBO
Document status	Final
Document version	V1.0
Date	30/06/2022
Authors	Christophe Guillotel-Nothmann (CNRS-IReMus) [Task leader] Jacopo de Berardinis (KCL) Thomas Bottini (CNRS-IReMus) Philippe Cathé (Sorbonne-IReMus) Enrico Daga (OU) Marilena Daquino (UNIBO) Achille Davy-Rigaux (CNRS-IReMus) Marco Gurrieri (CNRS-IReMus) Peter van Kranenburg (KNAW) Eleonora Marzi (UNIBO) James McDermott (NUIG) Albert Meroño Peñuela (KCL) Paul Mulholland (OU) Andrea Scharnhorst (KNAW) Rocco Tripodi (UNIBO)



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

## Project information

**Project Start Date:** 1<sup>st</sup> 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](mailto:valentina.presutti@unibo.it)

### Project Manager

**Marta Clementi**

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

Research division

E-mail: [marta.clementi3@unibo.it](mailto: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 (memoirs, reportages, news, biographies, reviews), different languages (English, Italian, French, Spanish, and German), and across centuries.

The overall goal of the project is to realise achieve 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

*Polifonia* is driven by ten pilot use cases which provide both a validation context and the input requirements to the other research and development work packages. The pilots are heterogeneous in terms of knowledge domains, e.g. bells heritage, popular Irish music, history of music in Bologna, music influence on children<sup>1</sup>. They involve interdisciplinary teams that bring different experience and methodological practices.

This report describes the efforts made in relation with the Technical Board (TB) (see deliverable 1.3<sup>2</sup>); the Data Management Plan (DMP) (see deliverables 7.1 and 7.4<sup>3</sup>); the working group (WG<sup>4</sup>) focused on the Web portal (see deliverable 1.9<sup>5</sup>); work package 2 (WP 2)<sup>6</sup> (see deliverables 2.1, 2.4 and 2.7<sup>7</sup>), WP 5 and the groups gathering the pilots and the work packages for the definition of a socio-technical Roadmap. This Roadmap aims to provide a methodological framework for the delivery of the 10 pilots. The initial version of the Roadmap (D1.1. produced at M6<sup>8</sup>) described the efforts made to gather the requirements and needs of the pilots on the basis of three tools: a) the Stories, b) the Maninpasta, and c) the Survey (SMS tools). This second version provides a definition of the methods, tools and strategies that regulate the socio-technical interactions and lead to an overall Validation framework.

After outlining the background to this deliverable and its status in Polifonia, the following four sections successively address the four main objectives of the Roadmap: 1. Coordinating collection of resources for use in the pilots, 2. Gathering requirements from internal and external adopters/stakeholders, 3. Setting objectives and challenges and mapping them from specific domains to the technology provider work packages, and 4. Monitoring validation within the pilot tasks. Six short appendixes make this report self-contained, including a glossary of terms, a description of the ten pilots, a description of *Polifonia* work packages' organisation and an updated version of the Survey.

## Document History

---

<sup>1</sup> The pilots are described in detail in appendix 2.

<sup>2</sup> Enrico Daga et al. *D1.3. Pilots development – collaborative methodology and tools*, Polifonia, 2021.

<sup>3</sup> Andrea Scharnhorst et al. *D7.1. First Data Management Plan*, Polifonia, 2021; Scharnhorst, Andrea et al. *D7.4. DMP webinar on dynamic data management*, Polifonia, 2021.

<sup>4</sup> Appendix 1 contains a list of acronyms with their definitions.

<sup>5</sup> Marilena Daquino et al., *D1.9. Polifonia Web portal – 1st version*, Polifonia, 2022.

<sup>6</sup> For a description of Polifonia's 8 work packages, see appendix 4.

<sup>7</sup> Albert Meroño-Peñuela et al. *D.2.1. Ontologies and knowledge graphs of music objects, patterns, and software package – 1st version*, Polifonia 2021; Albert Meroño-Peñuela et al. *D.2.4. Methods for interlinking knowledge graphs – 1st version*, Polifonia 2022; Albert Meroño-Peñuela et al. *D.2.7. Ontology testing and evaluation report; knowledge graph documentation and access interfaces – 1st version*, Polifonia, 2022.

<sup>8</sup> Christophe Guillotel-Nothmann et al. *D1.1. Roadmap and pilot requirements 1st version*, Polifonia, 2021.

<b>Version</b>	<b>Release date</b>	<b>Summary of changes</b>	<b>Author(s) - Institution</b>
V0.1	30/04/2022	First draft released	CNRS
V0.2	20/05/2022	Initial version submitted to internal reviewers	CNRS
V0.3	19/06/2022	Version updated according to the reviewers' remarks	CNRS
V1.0	30/06/2022	Final Version submitted to EU	UNIBO

## Table of contents

Project information .....	2
Project contacts.....	3
POLIFONIA consortium.....	3
Executive summary .....	5
Document History.....	5
Introduction .....	8
1. Coordinating the collection of resources, methods and tools used in the pilots.....	9
1.1 Goals, methods and strategies .....	9
1.2 Results .....	11
1.3 Lessons learned .....	20
2. Centralizing needs and requirements .....	21
2.1 Methods and strategies.....	22
2.2 Results .....	24
2.3 Lessons learned .....	28
3. Mapping objectives and challenges from pilot specific domains to the technology provider WPs .....	28
3.1 Goals, methods and strategies .....	28
3.2 Results .....	34
3.3 Lessons learned .....	35
4. Monitoring pilot validation .....	37
4.1 Goals, methods and strategies .....	37
4.2 Results .....	38
4.3 Lessons learned .....	42
FAIR protocol .....	43
Conclusion .....	44
References .....	46
Appendices .....	48
Appendix 1: Glossary.....	48
Appendix 2: List of the pilots.....	51
Appendix 3: WP1 participants, efforts, objectives, tasks and deliverables.....	53
Appendix 4: Work packages in Polifonia project.....	55
Appendix 5: Stories – updated template .....	58
Appendix 6: Updated Survey template .....	59



## Introduction

*Polifonia* is driven by 10 pilot use cases addressing preservation, management, studying musical heritage (MH), and interacting with it. Being real-world use cases – for example the Italian bell heritage, organs in the Netherlands, access to musical heritage for deaf people or the identification of recurring melodic patterns –, these pilots provide a means to validate *Polifonia*'s computational solutions, to strengthen the project's outreach, and to integrate new scenarios, early-adopters and stakeholders. They are developed by interdisciplinary teams of researchers from the humanities and social sciences, and from computer science.

This deliverable presents the second version of a “socio-technical Roadmap” which aims at providing a common framework to all pilots for:

1. Coordinating collection of resources for use in the pilots,
2. Gathering requirements from internal and external adopters/stakeholders,
3. Setting objectives and challenges and mapping them from specific domains to the technology provider work packages, and
4. Monitoring validation within the pilot tasks.

While the first version of the Roadmap, D1.1 released at M6<sup>9</sup>, adopted a bottom-up approach, the current version proposes a generalization of methods to share, interlink, articulate and validate resources, tools and results within a decentralized yet interconnected Ecosystem.

This generalization is based on an interdisciplinary dialogue between a) the musicologists, linguists, historians and curators working at the pilots and b) the computer scientists, designers and researchers in cognitive sciences involved in the project. The outcome is the result of a continuous interaction between WP1<sup>10</sup> – the driving force of *Polifonia* and place where the pilots happen – and the technology provider work packages (WP2-5).

An important vector of this dialogue are the “SMS-tools” – Stories, Maninpasta, and Survey –, which were presented in detail in the first version of the Roadmap. These tools have helped a) to constantly keep and update a centralized view of the information collected (Survey), b) to strengthen synergies through practical and concrete work between researchers from different disciplines (Maninpasta), and c) to collect the requests, needs and expectations of the pilots (Stories). As central pillars of the Roadmap, they have allowed both to identify and to promote common principles, methods and approaches among the pilots while preserving their specific needs and orientations.

The four sections of this deliverable relate to the four objectives outlined above and critically present how the tool's current state of development, methods and strategies contribute to the Roadmap. While the deliverable is organised according to the four objectives, the tools presented do not focus exclusively on one of them but cover several of them: the Survey, for example, contributes to both resource collection and validation monitoring. For methodological reasons, however, the tools will be presented in detail in the section in which they play a key role. Beyond the presentation of a set of tools, principles and methods that can be applied to the pilots, that regulate their interaction and

---

<sup>9</sup> See Christophe Guillotel-Nothmann *et al.* D1.1. Roadmap and pilot requirements 1st version, Polifonia, 2021.

<sup>10</sup> For a detailed description of WP1, see appendix 3.

that can be transferred to other use cases, the objective of this document is also to describe how the collaborative work took place and how the framework was shaped.

To facilitate the reading of this deliverable, a glossary (which provides a resolution of acronyms, short definitions and links) is included in Appendix 1. Appendix 2 gives a presentation of the pilots and their objectives. Appendix 3 summarizes the essential data of WP1, to which the Roadmap is attached and which constitutes the driving force of the project. Appendix 4 provides key data on the other Polifonia work packages. The last two Appendices 5 and 6 give an update of the pilot's Stories and of the Survey.

## 1. Coordinating the collection of resources, methods and tools used in the pilots

Three main devices are implemented to identify, organize and articulate the resources, methods and tools used by the pilots: a) the Ecosystem with its Rulebook, b) the Web portal and c) the Survey.

### 1.1 Goals, methods and strategies

#### 1.1.1 Ecosystem with Rulebook.

The Polifonia Ecosystem<sup>11</sup> – which is the collection of data, software, and documentation related to the project – allows the identification of all resources and tools created in Polifonia. The TB supervises the definition of the guidelines for collaborative development, encoded in a Rulebook, targeting consortium members as well as external stakeholders, following the *open science* principles. The TB reflects on the organization principles of the Ecosystem. This includes the provision of a set of descriptors through which developers can describe the components, allowing to support exploratory interfaces to navigate between them, tracing their evolution, and documenting the decisions made at both the technical and methodological level. As such, the TB actively participates in the effort to identify, organize and articulate resources, methods and tools that lie at the heart of the socio-technical Roadmap.

#### 1.1.2 Web portal.

The Web portal aims at collecting highly curated data produced by pilots and serving them, along with third-party data, via user-friendly interfaces. To produce a coherent and meaningful resource, an important task is the harmonization of content produced by pilot sources. Since the scope, the completeness and the granularity of data sources produced by data providers in the consortium may significantly vary, the primary goal of the Web portal is to define a sound narrative to connect music-related domains, expertise and research requirements, and to define the target of our expected impact.

On the one hand, a comprehensive strategy should be provided to target the broad audience of music passionate and professionals. On the other hand, preserving the diversity of approaches, methods and research interests should also be encouraged. Therefore, the design of the Web portal

---

<sup>11</sup> Browsable site: <https://polifonia-project.github.io/ecosystem/pages/pilots/>, Github repository: <https://github.com/polifonia-project/ecosystem>.

is articulated into two main research areas: a) design interfaces for lay people (i.e. without music skills and/or without any research task in mind) to enjoy music heritage collections via interactive Stories; b) design task-driven interfaces which are based on data-driven storytelling strategies, so as to accommodate scholars' research tasks.

To foster co-creation and co-design of the Web portal, we refined the methodological tools offered by the SMS tools, the Rulebook, the DMP and WP5 user interface (UI) / user experience (UX) design strategies<sup>12</sup>. In particular, we systematically analyzed Stories, Competency Questions (CQ)<sup>13</sup> and ontology requirements, to identify knowledge discovery patterns shared by pilots and organize content of the Web portal accordingly. Secondly, we had focus groups with scholars in the consortium and lay people (questionnaires with around 150 participants were collected) to iteratively collect expectations and potential frustration elements, as well as to validate our design choices. Lastly, we classified cognitive tasks underlying CQ and expectations, and we co-designed flexible, advanced data-driven interfaces (i.e. dashboards) to allow current and future (unexpected) research questions to be addressed via the Web portal interfaces without requiring us to revisit our work.

### 1.1.3 Survey.

The Survey<sup>14</sup> was developed in order to a) systematically and iteratively collect information about the pilots, b) allow an adjustment of the pilots through the examination of the answers given, and c) push their design through the questions asked. It consists of three main sections: the first one relates to the pilot's specific domain, the second provides technical information, and the third deals with validation and engagement issues<sup>15</sup>. The Survey's main objective is to create a map of the resources, methods and tools in relation with these three domains to articulate these components, to foster synergies and to identify common perspectives and particularities. The objective is not only to follow the evolution of the pilots through the different iterations<sup>16</sup>, but also to allow for progressive adjustments between them and to adapt the Survey to the problems encountered and the questions that arise at different moments of the project.

Compared to the first iteration of the Survey, described in deliverable 1.1, several adjustments have been made. In the domain-specific part, questions have been added about the pilots' objectives, the results obtained and the progress made towards the full achievement of the objectives. In the technical part, questions were added about the expectations of the pilots towards the technology providers (WP2-5), the collaborative space of WP1, and towards the Web portal (presentation of resources, type of knowledge made available, types of interactions, link with the pilot-specific demonstrator, added value of the Web portal). The "engagement and validation" part (former "socio-pedagogical part") has been extensively modified, too. Questions were added about the identification of the pilot's partner institutions (beyond the coordinating institution) and an inventory

---

<sup>12</sup> For details, see Marilena Daquino *et al.*, *D1.9. Polifonia Web portal – 1st version*, Polifonia, 2022.

<sup>13</sup> See section 2 below about these concepts.

<sup>14</sup> The Survey is available on github: <https://github.com/polifonia-project/survey>.

<sup>15</sup> The questions from the Survey, in its current version, are reproduced in Appendix 6. The answers given by the pilots are available on the Polifonia Github: <https://github.com/polifonia-project/Survey/blob/main/Survey.xlsx>.

<sup>16</sup> Iterations took place at the following dates : 07.05.2021, 14.05.2021, 15.11.2021 and 13.06.2022. See above in this document, section 1, table 1.

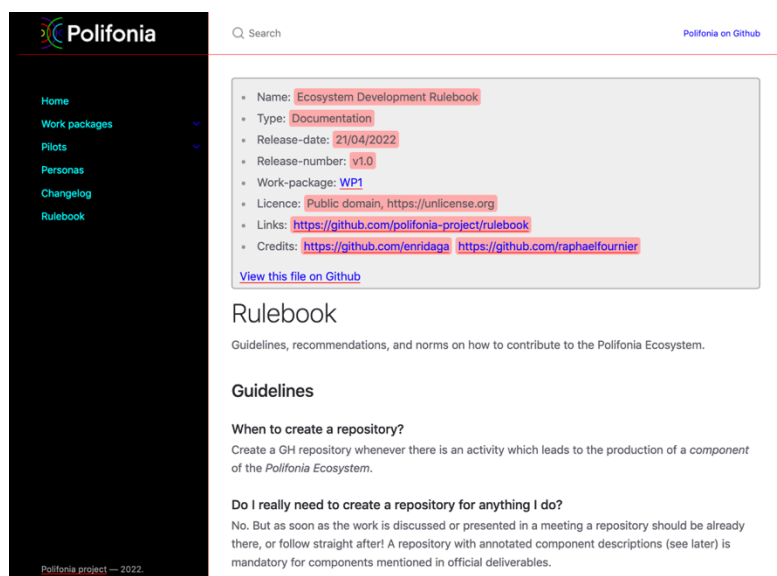
of how the feedback from users/experts has been considered has been requested. The pilots are also asked about the demonstrators that will be made available in the coming months. Finally, questions related to pilot monitoring and validation were added/merged.

The Survey was originally an Excel spreadsheet that was progressively updated by the pilot leaders. Our goal for the remainder of the project is to transform the Survey into a knowledge graph about the pilots: the pilot knowledge graph (PiKG). The motivation behind this is to identify both the components that are shared between the pilots and the pilot-specific concepts. (e.g. formats, standards, collections, licenses, methods, epistemological positions, scientific questions, validation devices, etc.). As a complement to the Polifonia Ontology Network (PON), which aims at interconnecting the pilots through the entities and properties they use, the interconnection made by the Survey is at a meta-level. The Survey provides a framework to identify and document connections – methodological choices, scientific assumptions, ways of engagement, validation protocols – that cannot be deduced by analyzing common entities/properties. The different iterations of the Survey also show how the pilot evolved over time, how methodological convergences appeared, how they refocused, etc.

## 1.2 Results

### 1.2.1 Ecosystem.

The methodology for the collaborative development that led to the design of the Ecosystem was already presented in D1.3<sup>17</sup>. Since then, the TB designed a set of guidelines, recommendations and standards to regulate the way consortium members (and stakeholders) can contribute to the Ecosystem. These results are presented in the GitHub Polifonia Organisation, specifically, in the repository Rulebook<sup>18</sup>.



Polifonia

Search

Polifonia on GitHub

Name:	Ecosystem Development Rulebook
Type:	Documentation
Release-date:	21/04/2022
Release-number:	v1.0
Work-package:	WP1
Licence:	Public domain, <a href="https://unlicense.org">https://unlicense.org</a>
Links:	<a href="https://github.com/polifonia-project/rulebook">https://github.com/polifonia-project/rulebook</a>
Credits:	<a href="https://github.com/enridaga">https://github.com/enridaga</a> <a href="https://github.com/raphaelfournier">https://github.com/raphaelfournier</a>

[View this file on Github](#)

### Rulebook

Guidelines, recommendations, and norms on how to contribute to the Polifonia Ecosystem.

#### Guidelines

**When to create a repository?**  
Create a GH repository whenever there is an activity which leads to the production of a *component* of the Polifonia Ecosystem.

**Do I really need to create a repository for anything I do?**  
No. But as soon as the work is discussed or presented in a meeting a repository should be already there, or follow straight after! A repository with annotated component descriptions (see later) is mandatory for components mentioned in official deliverables.

### Example 1: Ecosystem, Rulebook.

<sup>17</sup> Enrico Daga et al. D1.3. *Pilots development – collaborative methodology and tools*, Polifonia, 2021.

<sup>18</sup> <https://github.com/polifonia-project/Rulebook>.

The Rulebook includes an *annotation schema* for Ecosystem components that allows to annotate each component with a rich metadata set (example 1). It defines when to create a repository (i.e. for any component of the Ecosystem), how to associate it with a "champion" (i.e. the repository leader), how to structure it and how to enrich it with metadata that: identifies the component within the Ecosystem, defines it by keywords, links it to the different WPs and pilots, assigns it an order of priority, and classifies it according to a typology. As an example, the Amy#2\_OrganBuilders component (example 2) is a "Story" component (i.e. a narrative that sets out a certain problem to be explored)<sup>19</sup>, related to WP2 and to the Organs pilot. This Story's Persona (i.e. the protagonist of this Story) is "Amy" and is about "durch organs", "instruments", traditions".

main
stories / Amy: Organologist / Amy#2\_OrganBuilders.md
Go to file

enridaga See polifonia-project/rulebook#26
Latest commit cb2a847 on Mar 11
History

2 contributors

48 lines (38 sloc) | 1.86 KB
Raw
Blame

	persona	keywords	work-package	pilot	priority	type	story type
rs	Amy	dutch organs instruments tradition	WP2	ORGANS		Story	expert driven

### Amy#2\_OrganBuilders

#### Goal

Amy wants to assess the developments of organ builders.

#### Scenario

This research includes looking into which organs an organ builder worked on and how their projects developed over time. This development can range from an increase of the size of the organ, to locality of their projects, or to the type or prestige of the project. The history section of the 15-part organ encyclopaedia is mostly used for this as this specifies year and changes that are made to the organs, but this obviously takes a lot of time and effort to research.

### Example 2: Amy2\_OrgannBuilders as an Ecosystem component.

Such methodological practices contribute to make the whole set of components explorable through an Ecosystem website. Based on the metadata provided, the components are indexed in the Ecosystem and can be navigated according to the structural relationships defined. This approach allows the publication of components in real time, as Polifonia progresses and is an important methodological outcome that could be published itself with a view to being made reusable and adaptable to other projects. The first version of the Polifonia Ecosystem was released in April 2022 (Milestone 2) - v1.0<sup>20</sup>. New releases are expected in line with the software deliverables and the other milestones of the Polifonia Project.

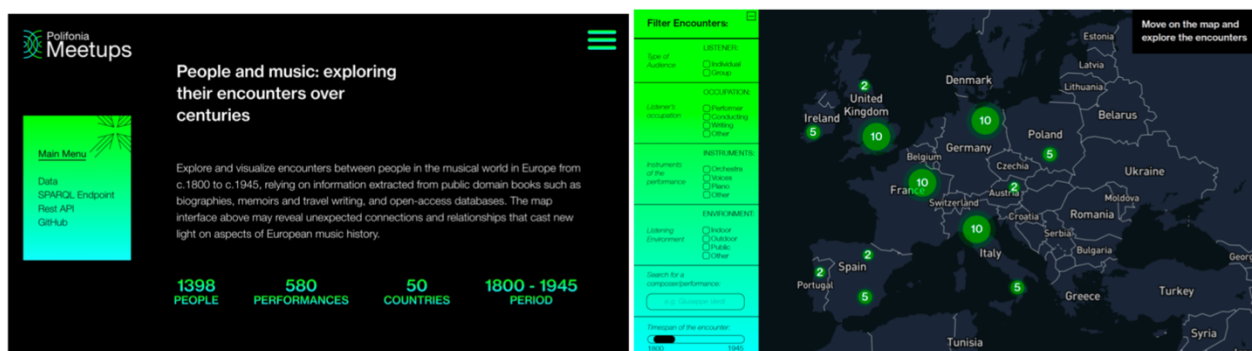
<sup>19</sup> The „Story“ concept will be explained in detail in section 2 of this document.

<sup>20</sup> <https://polifonia-project.github.io/ecosystem/>.

### 1.2.2 Web portal.

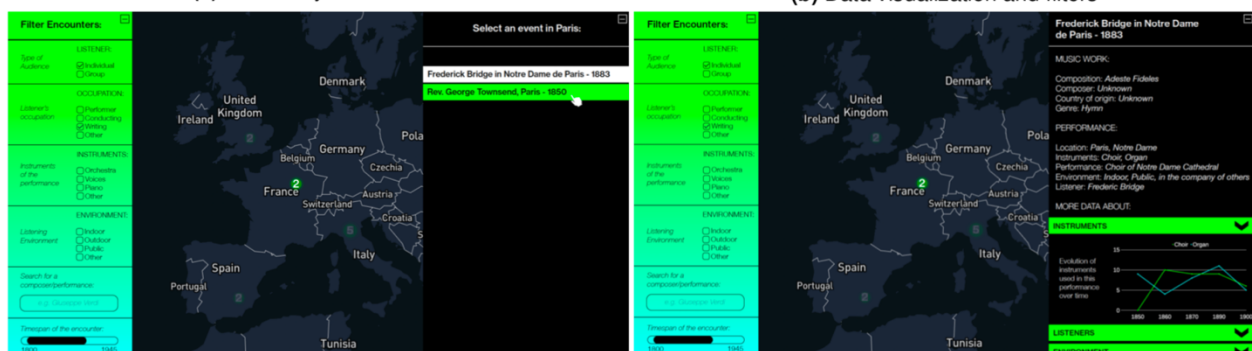
While the design of interfaces for lay people had limited appeal to scholars in the consortium (and will therefore be subject of further study in the next phase of the project), the design of data-driven dashboard interfaces led us to grasp a comprehensive outlook on research methods (i.e. qualitative and quantitative) and expectations of scholars with different interests and levels of expertise.

In particular, we were able to classify research methodologies into three schemata, in turn represented by three dashboard templates. Templates address different (incrementally difficult) types of analysis and interaction patterns, namely: statistics-based and qualitative argumentation, geographic-based interactive exploration, charts manipulation for datasets integration and temporal analysis. These templates have been developed on the basis of the Meetups, Musicbo, Bells, Child, Tunes and Organs pilots whose only interface will be the Web portal, while the pilots with an independent interface (Interlink, Facets, Tonalities) have not been taken into account at this stage<sup>21</sup>. In the case of Meetups, which explores the encounters between musicians between 1800 and 1945, the templates provide general information on the number of people, performances and countries considered (example 3a). They allow geographical visualization of the data based on filters (example 3b), give general information on individual data points (example 3c) and provide tools for granular exploratory visualization of data points, for example on instruments used, listeners and socio-cultural environments (example 3d).



(a) Data story introduction

(b) Data visualization and filters



(c) Data point insights

(d) Exploratory visualizations related to data points

**Example 3: Web portal, MEETUPS prototype, geographic-based interactive exploration<sup>22</sup>.**

<sup>21</sup> The Access pilot has not been considered either.

<sup>22</sup> See Marilena Daquino *et al.*, *D1.9. Polifonia Web portal – 1st version*, Polifonia, 2022, section 4.4.1.



While more interaction patterns and types of analysis could be relevant in the future, we believe that we achieved a reasonable coverage of requirements, which will allow us to share meaningful insights from pilot datasets that accommodate most desiderata<sup>23</sup>.

### 1.2.3 Survey.

The Survey provides detailed insights into the pilots' objectives, resources, methods, tools and progress. This tool is a central entry point to the components related to the pilots (including the Personas, Stories, and Competency questions (see below section 2)) and does not only allow to explore the individual pilots but also to identify the interconnections between them. As such, the Survey has proven to be a valuable tool for the coordination of the pilots and for their articulation within Polifonia.

The information was collected in four iterations of the Survey, listed in bold in table 1. The first iteration was carried out during the first weeks of Polifonia and aimed both at obtaining initial data and adapting/refining the questions. Iterations 2 (17/05/2021) and 4 (13/11/2021) are situated in the direct context of the two versions of the Roadmap which are central milestones in the update of pilot information. Iteration 3 (15/11/2021) follows on from the 2nd Project Meeting (Bologna, 13-15/10/2021) which allowed the pilots to refine their objectives and methodological choices. Independently of the completion of the Roadmap, it is planned to continue updating the Survey by correlating it with other important future moments for the pilots: the fifth Project Meeting (Paris, 22-25/09/2022) the pilot's mid-term evaluation (M24) and their final evaluation at M40.

Commit	Date
<b><a href="#">Survey initial publication (March 2021)</a></b>	<b>07/05/2021</b>
<b><a href="#">Update Survey after 2nd iteration (14 05 2021)</a></b>	<b>17/05/2021</b>
<a href="#">Reorganized formats and standards in technical part</a>	17/06/2021
<a href="#">Merged with DMP Annex 1, col F and G</a>	17/06/2021
<a href="#">Minor format update</a>	17/06/2021
<b><a href="#">New iteration: validation protocols and general update</a></b>	<b>15/11/2021</b>
<a href="#">New iteration: validation protocols and general update</a>	19/05/2022
<a href="#">Objective questions and validation questions updated</a>	13/06/2022
<b><a href="#">Survey 4th iteration : validation and stakeholders</a></b>	<b>13/06/2022</b>

**Table 1: Survey commits on GitHub.**

**a) Domain-specific part.** The Survey's first part provides granular information on the pilot's objectives, results, potential innovations, input data, working hypotheses, knowledge produced, tools, concepts, controlled vocabularies (created or used), disciplines involved, etc. Since the first version of the Roadmap, the pilots have not only documented their objectives, main innovations and contribution to the state of the art, they have also detailed their achievements until now and estimated the progress made towards the realization of their overall objectives (see table 2). The goal of this part is thus to position the pilots in the current scientific landscape, to identify their ability to answer questions of central importance, and to assess their interest for the community of

<sup>23</sup> As confirmed by the evaluation with scholars described in D1.9. See Marilena Daquino *et al.*, *D1.9. Polifonia Web portal – 1st version*, Polifonia, 2022.

researchers, students, professionals, industries and amateurs who are interested in cultural heritage. Furthermore, the information collected allows us to monitor the pilots' progress: the pilots have already completed some important tasks – the constitution of their corpora (Musicbo), the creation of resources for the interconnection of collections (Interlink), the design of a theoretical model for the indexing of scores (Facets), etc. – and are all planning in line with the project's duration.

Contribution to the state of art	What are the pilot's main innovations ? / How do the methods and results contribute to the state of the art?
MUSICBO	The main innovations of the pilot are: 1 creation of a corpus dedicated to the role that music has played for the city of Bologna; 2 extraction of new knowledge relating to the links between music and the history of the city has remained latent until now
TONALITIES	Tonalities aims to compare different (sometimes controversial) music analyses that MUST be referred to specific theoretical frameworks. In this way, the pilot innovates by contextualising theoretical and analytical information. The open linked data paradigm allows to articulate different large digital collections of notated music and to explore them through a quantitative-qualitative approach. By modelling different theories and applying them to musical works in a dedicated interface combining machine learning and human annotations, we can analyse larger score collections and deepen our understanding of them.
CHILD	Cultural studies in music: a systematic study of music and childhood in the reference period is still to be done. State of art knowledge extraction methods can partly be used (eg. named entity recognition) but new research is needed to support factual information extraction such as event extraction and characterization for the domain of interest.
MEETUPS	Contributes new tools and methods for building a KG of historical meetups. The resources contributes to historical musicology research.
ACCESS	By contributing new tools, frameworks, perspectives and an understanding of their effects.
BELLS	Building a formalized knowledge about the phenomenon, which is non available at the moment
INTERLINK	Largest collection available of MIDI files harvested from the Web, cleaned and aligned with the Million Song Dataset; Largest collection of open and high-quality chord annotations in different formats and notations with timing information. Tools for dataset enrichment and linking.
FACETS	The SotA on search-engine for musical scores is rudimentary, and was mostly started by our team. We aim at extending previous preliminary theoretical work and produce working implementations.
ORGANS	In the first place, the results will facilitate the research that is needed for restoration and maintenance of the organs. Results on trends and longitudinal developments will contribute to better understanding of the individual instruments, and provide a better understanding of history of pipe organ building.
TUNES	Methods for analysing music (melodies) at a higher abstraction level than single notes are lacking in Music Information Retrieval. We will contribute to that.

Goals and objectives	What is the progress towards the pilot's objectives (please indicate %)
MUSICBO	*Design and collection of corpus 100% *Modeling extraction of textual information 20%
TONALITIES	*Classify musical works into modal-tonal categories on the basis of an in-depth understanding of their inner organisation 10% *Develop and apply at least 5 different theoretical models to identify how the work's modal-tonal interpretation changes through the lense of a given theoretical framework; 40% *Compare modal-tonal attributions of a musical work; 10% *Look "inside" modal-tonal theories and compare them; 100% *Provide an argued, documented and authored modal-tonal analysis on the basis of the theoretical and analytical observations made. *Develop a semantic model for generic score collaborative annotation; 20% *Design & develop a Web interface and a Web service for collaborative score annotation; 20%
CHILD	*Develop a music history study; 30% *Discovery and collection of resources from books; 50% *Automatic curation of documentary evidence; 30% *UI for visualisation, exploration, and sense-making; 0% *Knowledge Graph of Music and Childhood; 20%
MEETUPS	User story: 100% Requirements: 50% (UI requirements missing) Ontology: 50% (first version, to be further developed) Knowledge Graph: 30% (first version focused on 1000 bios) Study in the music domain: 20% (focus on Edward Elgar)
ACCESS	
BELLS	
INTERLINK	Creation of new resources to facilitate and enable the linking operations (80%); Discovery of links at the entity level (10%); Discovery of links at the musical content level (33%);
FACETS	*Theoretical model for indexing scores efficiently: 90%. *Features of music to extract: 50%. *Faceted search: 0%. *Software distribution: 30%
ORGANS	
TUNES	

**Table 2: Survey, domain specific part, main innovations, progress towards the pilot's objectives.**



Another field of work concerns the licenses and copyright of the sources on which the pilots are based. If the cultural heritage is in the public domain, the facsimiles, transcriptions, electronic editions – for example an edition of the works by Palestrina or a madrigal by Tasso – on which the pilots might focus, may be under copyright. The Survey allowed to collaboratively contribute to the identification of the licenses and copyrights of the collections taken into account and to design a contact scheme according to the different solutions encountered. The aim is to identify, in close connection with the DMP, the challenges that could arise in order to achieve the objective of making reusable the pilot's outputs, including for commercial purposes by industries.

**b) Technical part.** The technical section reviews the formats and standards of the data and metadata used. It identifies the controlled vocabularies, thesauri and ontologies, the interoperability/alignments envisioned, and the tools and algorithms developed. This part also highlights the (envisioned) overlaps between the pilots, the expectations of the pilots towards the technology providers, and the needs addressed at the WP1 level. Finally, this part questions the pilots on the application they will release and on their expectations regarding the Web portal.

Since the first version of the Roadmap, the pilots have clarified the areas of shared work, for example between Musicbo, Meetups and Child; or Tunes, Facets, Interlink and Tonalities (see table 3). These interconnections crystallizing through the Survey and being exploited by the pilots – Interlink and Tunes have an active and ongoing collaboration, for example – play a central role in coordinating resources. They contribute to the goal not to juxtapose Polifonia's studies but to articulate them in a rich network that reconciles complementary perspectives.

In the past months, the pilots have also clarified their expectations towards the technology provider WPs (i.e. WP2-5) (see table 3). The expectations towards WP2 are mainly related to ontology design (Tonalities), methods and tools for data extraction from knowledge graphs (Meetups) or conversely the conversion of pre-existing data (textual and/or musical) into linked data (Interlink, Organs, Tunes). The pilots also aim to build on the expertise of WP2 for the extension and harmonization of metadata schemas (Bells), the design of algorithms for linking musical scores and metadata (Interlink) and the representation of knowledge graphs (Organs). The collaborations with WP3 are mainly focused on the extraction of musical patterns (Tunes), partly based on neural networks (Tonalities), as well as on the definition of musical patterns and contents (Facets). Conversely, WP4 is also involved in knowledge extraction from textual data (Musicbo, Child, Meetups, Organs). Finally, WP5 is expected to contribute to the design of interfaces (Tonalities, Child, Meetups) and to the design of general dashboards (Tunes). The relationships between pilots and technology providers, uncovered by the Survey, have played an important role in a) the collection of resources used in the pilots, b) the centralization of expectations (see section 2 below) and c) the mapping of disciplinary needs to technologies. In particular, the associations between pilots and WPs have led to weekly working groups which are now a driving force for the project (see section 3 below).

	Are there planned overlaps (already defined or envisioned) between the pilot's resources and other resources used within Polifonia?		What do you expect in terms of technologies from the technical provider WPs?
MUSICBO	*The sources of MusicBo and MEETUPS are complementary and can mutually enrich the two pilots.	MUSICBO	WP4 -> Knowledge extraction components; WP5 -> UI design
TONALITIES		TONALITIES	*WP2: Help in ontology design and alignment. Strategies for ontology-based reasoning *WP3: Help in the design/training of a neural network for rate analysis *WP5: Contribution to the design of the collaborative annotation interface.
CHILD	*The Polifonia Corpus (probably)	CHILD	*Knowledge extraction research performed in WP4 *UI research developed in WP5 for the exploration/curation of accounts *UI design
MEETUPS	*Sources from MusicBo may contribute data about Bologna and the CHILD pilot may include texts that could be used to extract encounters as well. *Shared interest in the theme of meetings between musicians: therefore it is possible to share a portion of the corpus that deals with the Meetups that took place in Bologna	MEETUPS	*WP2 to provide methods and tools for extracting content from existing knowledge graphs (Wikidata/DBpedia/...) publishing linked data *WP4 to provide methods to extract encounters from Wikipedia and digitized texts *WP5 -> UI design
ACCESS	*Possibly in WP8	ACCESS	We are our own technical provider, though we will collaborate with others within and outside the project, as applicable .
BELLS		BELLS	*Data enrichment *Possible extension and harmonisation of metadata schemes *Data sources hosted by content service providers; *Map hosted by Polifonia portal *Publish/enrich LOD
INTERLINK	*Yes (e.g. MIDI Linked Data cloud)	INTERLINK	*Linked Data conversion of MIDI and other notations to RDF, *Ontologies for these notations *Linking algorithms between notations and metadata (e.g. related texts, events)
FACETS	*Yes, definitely with TONALITIES	FACETS	*Guidelines so that we may easily integrate in the Polifonia ecosystem. *WP3 : pattern definitions / dimensions of musical content relevant for search
ORGANS		ORGANS	*NLP: tailored named entity recognition, techniques to infer relations between the entities from the text. *KG: Support for extracting a knowledge graph from the data. WP2: knowledge graph representation *WP4: Natural Language Processing
TUNES	*There is envisioned overlap with TONALITIES and INTERLINK. *Also with FACETS.	TUNES	* WP3: technology to extract patterns from the musical contents. * WP2: technology to extract a knowledge graph from these extracted patterns. *Generic dashboard; *Convert dataset(s) to linked data; RDM: enriching existing database (?) similar to Bells; application possible in Portal (?); see BELLS on SLA's

**Table 3: Survey, technical part, planned overlaps between the pilots and expectations towards the technology providing WPs (WP2-5)**

A new field compared to the first version of the Roadmap concerns the description of the application that will be published by the pilots and their relationship to the Web portal (see table 4). The objective is to lead the pilots to sketch as soon as possible what will be pilot specific and part of their individual output and. A complementary goal is to exploit the canalizing potential of the Web portal as a shared environment to aggregate, articulate and interconnect the individual resources. The Survey thus aims to reinforce the transversal dimension of the project and to generate added value through the identification/documentation of inter-team collaboration.

The answers given show that while most pilots suppose an independent application (Musicbo, Tonalities, Access, Bells, Organs, Tunes, Facets) some of them do not exclude relying on the Web

portal only (Child, Meetups, Interlink). Independently, all the pilots agree on the need for a joined (yet not exclusive) presentation through the Web portal. The portal thus not only contributes to the coordination of resources. It also has a channeling effect on the technical and even methodological level, through the presentation of the pilots' resources.

	What is the expected application to be released for this pilot?		What type of interaction do you expect should be supported by the Web portal?
MUSICBO	A knowledge map of Bologna, which shows semantic relations about music in some specific points of the city, as places, public or private building, churches and so on.	MUSICBO	Browsing and querying of the Pilot corpus
TONALITIES	Provide an interface for annotation and analytical argumentation. This interface would <ul style="list-style-type: none"> <li>- Classify musical works into modal-tonal categories on the basis of an in-depth understanding of their inner organisation;</li> <li>- Apply different theoretical models to identify how the work's modal-tonal interpretation changes through the lense of a given theoretical framework;</li> <li>- Confront modal-tonal attributions of a musical work;</li> <li>- Look "inside" modal-tonal theories and compare them;</li> <li>- Provide an argued, documented and authored modal-tonal analysis on the basis of the theoretical and analytical observations made.</li> </ul>	TONALITIES	Exploration of the information on the works, navigation in the corpus, discovery of the available analyses, navigation towards the specific display resources of TONALITIES.
CHILD	A) A knowledge acquisition UI for listening experiences, linked to the LED B) A knowledge graph of listening experiences (LED+updates focusing on CHILD) C) A UI for thematic exploration of LEs, focusing on Childhood (->WPS)	CHILD	Exploration of the information on the works, navigation in the corpus, discovery of the available analyses, navigation towards the specific display resources of CHILD.
MEETUPS	Website to allow the exploration of meetups through timeline and map Knowledge graph of musical meetups	MEETUPS	Exploration of the information on the works, navigation in the corpus, discovery of the available analyses, navigation towards the specific display resources of MEETUPS.
ACCESS	UI for Poly query UI for Poly Browse UI for Harmony Browse	ACCESS	For this particular pilot, the portal will be not be a major focus of interaction.
BELLS	<ul style="list-style-type: none"> <li>- Map, to locate the bells</li> <li>- Facet research</li> <li>- Time period in layers</li> <li>- Customize the criteria/facets</li> <li>- Natural language question answering over knowledge graphs - queries in natural language translated in SPARQL</li> <li>- issue: can the data be enriched by users ?</li> <li>- issue: can we guide the user ?</li> </ul>	BELLS	
INTERLINK	identify similarities in heterogeneous corpora	INTERLINK	Note-by-note navigation, score representation, playable on demand, browse and search specific measures
FACETS	Provide a faceted search engine for large collections of musical documents	FACETS	Search engine, with demo queries provided so the user can grasp the power of the tool.
ORGANS	The primary goal is to populate an existing database, developed by the NivO (National Institute for Organ Art) with data extracted from the raw text of the organ encyclopedia. The NivO already prepared a web-interface to the database: pipeorgan.nl  Additionally a dashboard which allows exploration of the knowledge graph that is extracted from the full text of the encyclopedia. Various visualizations and querying methods could be included.	ORGANS	No clear vision yet
TUNES	<ul style="list-style-type: none"> <li>- collection of tools</li> <li>- issue: who will build this application ?</li> <li>- issue: how do we represent melodic contents as KG ?</li> </ul>	TUNES	Maybe some sort of exploration tool.

**Table 4: Survey, technical part, expected applications and types of interactions supported by the Web portal.**

**c) Engagement and validation.** The "engagement and validation" part has been subject to the most extensive adaptations. The answers related to user feedback show that the pilots are aware of the need to involve, from an early stage, external experts (see below section 2.2). The other area with the most significant extensions in this section concerns validation<sup>24</sup>. The pilots were asked to specify their success criteria, to identify validation methods, to define a validation setup (see table 5), and to establish KPIs/metrics. The Survey shows that the pilots see success criteria in the quality of the data (Meetups), the networking of information from different sources and/or formats (Tonalities, Interlink, Organs, Tunes) and the speed and accuracy of task execution (e.g. counting

<sup>24</sup> While this section gives an overview of the information on validation provided by the pilots, the actual Validation framework will be discussed in section 4.



musical patterns in large corpora or identifying textual resources) (Tonalities, Child, Interlink, Facets, Organs). At the same time, the pilots emphasize the need to create innovative and ergonomic tools (Musicbo, Tonalities, Bells, Facets, Organs), made available to a wide audience and involving new forms of engagement and accessibility (Access).

The validation criteria that emerge from the Survey are strongly oriented towards validation by experts and/or target audiences. This concerns in particular the pilot's objectives, the input data, interfaces and results. The KPIs reflect the importance given to the target users satisfaction (in terms of objectives, methods, interfaces and results), the number of scholarly activities supported, and the use of the tools in educational situations or by cultural heritage institutions. These indicators are complemented by quantitative and statistical approaches that focus on the measurement of responses based on f-scores, the number of items processed, the size of KGs or the speed of query processing.

	What will be the validation setup?
MUSICBO	Validation of a case study (one buildings/place) by a musicologist.
TONALITIES	<ul style="list-style-type: none"> <li>* Component 1 : Goals and objectives: Do the objectives of the pilot correspond to real musicological needs? Do they take into account suggestions made by musicological colleagues?</li> <li>* Component 2: Analytical results: Does the theoretical model allow for the correct inference of modal attributions? - Use a test corpus (c. 150 works) and check whether the attributions made by the theorists can be inferred from the models. Use the justification for the modal inference produced by the reasoner as metrics ?</li> <li>*Component 3: Data processing: Is the analytical information provided by user(s) or computed by algorithms correctly stored and retrievable ? - Use a test corpus of multi-layered analyses (c. 15 works which have more than 1 analysis) and check if the information layers are stored correctly</li> <li>*Component 4: User engagement / ergonomics: Does the music annotation environment meet usability requirements ? Can de CQ be answered with the interface and tools provided ?</li> </ul>
CHILD	<ul style="list-style-type: none"> <li>*Quantitative: ability of the system to produce accurate recommendations to curators</li> <li>*User study on usability of thematic explorer</li> <li>*User study on assisted formulation of research hypotheses</li> </ul>
MEETUPS	<ul style="list-style-type: none"> <li>*User study with domain experts on how the tool can support scholarship</li> <li>*User study on the quality of the meetups KB</li> </ul>
ACCESS	<ul style="list-style-type: none"> <li>* Usability, accuracy, user feedback</li> <li>* Quantitative: ability of the UIs to pass accurate well-formed query to back ends</li> </ul>
BELLS	<ul style="list-style-type: none"> <li>* Validate the methodology : 10 case-studies to implement a methodology and good practices to build a complex knowledge system (ex: links between bells and sounds, bells and bell tower, performing events, archival documentation, playing techniques).</li> <li>* N19- interviews for validating the interface</li> <li>* Check that the corpus is cleaned, consistent through experts</li> </ul>
INTERLINK	<ul style="list-style-type: none"> <li>*Quantitative evaluation of links according to clustering metrics and ground-truth data;</li> <li>*Quantitative evaluation of music similarities based on MIR benchmark tasks;</li> <li>*Qualitative evaluation of both, through the involvement of collection curators/experts and musicologists respectively.</li> </ul>
FACETS	<ul style="list-style-type: none"> <li>*Technical aspects : Information Retrieval metrics (precision, recall, accuracy), computational time to execute queries. Identification of hard/specific score problems for retrieval.</li> <li>*Evaluation setup with a survey from experts, for the UI part.</li> </ul>
ORGANS	<ul style="list-style-type: none"> <li>*Component 1. Measurement how much of the information from the encyclopedia has been transferred correctly to the NivO database.</li> <li>Component 2. We will perform a user test, comparing the time certain queries take manually versus with the dashboard.</li> </ul>
TUNES	<ul style="list-style-type: none"> <li>*Output: Validate the output by using the existing manual annotation of tune origins as ground truth.</li> <li>*Interface: User study with domain experts on how the tool can support scholarship.</li> </ul>

**Table 5: Survey, Engagement and validation part, validation setup.**

An important challenge for the pilots is to establish individual validation protocols based on the identified success criteria, methods and KPIs. At this stage, these protocols combine between one and five components that cover the areas of objectives, methods, data processing, usability, and

results. All this information on validation has been decisive in the implementation of a global Validation framework for Polifonia which will be described in section 4.

## 1.3 Lessons learned

### 1.3.1 Ecosystem.

The first release of the Polifonia Ecosystem was the first time the consortium experimented with the collaborative methodology and, therefore, collected feedback from project partners. Such feedback was informally collected during TB meetings. The general approach has proven effective in that it allowed to decouple the effort required in developing the software output of the specific research activities without enforcing strong dependencies between developers and their related workflow. In addition, the common annotation schema seemed to add a sustainable overhead to the effort required to prepare the Ecosystem release. The majority of the components released in the first version were successfully composed into the Ecosystem Website and required minimal effort in metadata curation. The success of this methodology brought to the attention new opportunities. Specifically, we plan to extend the annotation schema to integrate metadata relevant to describe the pilot Validation framework (see below section 4) and formally describe the licensing and terms of use of the components. Finally, we plan to enrich the description of *how* the components interact with each other.

### 1.3.2 Web portal.

The co-design of interfaces and methods for promoting results of pilot-driven research has proven to be effective in a) preventing the lack of clear goals in the production of data sources, highlighting the importance of final users, and b) defining impact-driven application studies, such as designing actual data stories that exemplify the added value of produced data. Moreover, the continuous interaction with partners on the preliminary analysis of Stories and CQ allowed us to increase self-awareness of the pilots' status, revising expectations and objectives.

### 1.3.3 Survey.

The Survey has proven to be effective in coordinating the resources, methods and tools used in the pilots. In conjunction with the Web portal and the Ecosystem, it has been valuable for setting up a methodological framework within which the pilots can evolve. Its role is not restricted to collecting information, but also to giving impulses and steering methodological work. The Survey stimulates the pilots by asking them to specify their methodology, their objectives and the way they intend to involve experts. It also inspired the DMP, which worked on its elaboration and, based on this, identified more precisely some of the pilots' needs (e.g. fine-tuned analysis of licenses and copyright of external resources). Indirectly the Survey also stimulated the development of complementary tools – i.e. the Stories – to capture, from a discursive and narrative angle, the individual orientations, needs and requests of the pilots (see below section 2). The Web portal both stimulated and benefits from the responses obtained for the identification of the pilots' needs and can potentially use the information collected for the publication of pilot “identity cards”. Finally, the Survey contributes to the metadata of the Ecosystem in the context of the Validation framework (see above in this section and below section 4). As such, this tool contributes to stimulating the coordination of resources, methods and tools.

Another contribution lies in the ability of the Survey to network the answers provided and thus to create an interconnected map of methods, tools and resources. The ongoing conversion of the Survey into a knowledge graph will make this map even more browsable. In particular, it will allow for a more systematic exploration of the conceptual interconnections that arise between the pilots while providing access to the pilot's DNA by delivering highly granular information about them.

The Survey could in principle be completed with this final version of the Roadmap. However, it is planned to continue feeding and adapting it until the end of the project. On the one hand, our plan is to document the ongoing updating and validation process of the pilots (at M24 and M40). On the other hand, the Survey, once transformed in a knowledge graph, will provide an interconnected and permanent entry point to all the information and resources relating to the pilots. This entry point will allow to check the status of the pilots and their related resources (data, KGs, ontologies, software bricks, etc.).

Beyond, the questions and answers of the Survey will be explored and updated in the context of other WPs and deliverables. The second version of the "Methods for interlinking knowledge graphs" (D2.5 M36)<sup>25</sup> may exploit the information available to discover links between music objects and contexts in Polifonia. The second version of the "Plan for Exploitation, Dissemination and Communication of Results"<sup>26</sup> will be able to use the Survey to produce both a quantitative and qualitative report on Polifonia's pilot-related results achieved compared to the objectives set. Finally, the technical questions of the Survey will be taken up in the next two iterations of the DMP (D7.2<sup>27</sup> and D7.3<sup>28</sup>) which will focus on FAIR principles. In this context, the Survey will be an important cornerstone for documenting the pilot's individual relation to FAIRness.

Despite all these contributions and outcomes, it must be kept in mind that the Survey is based on a mainly static interaction with the pilots: they are invited to answer questions that have been formulated beforehand and that are only adapted during new iterations. While the tool is therefore suitable for the methodological generalization that this deliverable aims to produce, it has to be supplemented by other tools that are more adaptable to the individual pilots and their narratives.

## 2. Centralizing needs and requirements

The centralization of needs and requirements happens at two levels within the pilots. On the one hand, a central issue is to actively involve external stakeholders as early as possible in the design and validation process. The objective here is to ensure that the pilots reflect real expectations of the communities related to cultural heritage, that they are used by these communities, that they can be adapted to similar problems, and that they will survive the project itself. On the other hand, at

---

<sup>25</sup> D2.4 aims at providing a technical description of the method to discover links between music objects and contexts in Polifonia and external datasets, at the schema and data levels.

<sup>26</sup> D.6.2 contains a first quantitative and qualitative report on planned and completed communication and dissemination activities, success measures against the stated goals, and recommended modifications to the original strategy. Also, the initial Exploitation and Business Plan outlines the exploitation activities planned for during the project and an initial exploitation and business plan based on initial feedback from industry experts.

<sup>27</sup> D7.2 consists of an updated version of a Data Management Plan (DMP) to support the management of the outputs of the project during the development of the project, and after the project has ended.

<sup>28</sup> D7.3 is the Final version of a Data Management Plan (DMP). Its objective is to support the management of the outputs of the project during the development of the project, and after the project has ended.

the internal level, the objective is to identify the technological needs of the pilots. More precisely, this involves a) defining the structure of the Polifonia Ontology Network (PON) based on the needs of the pilots and b) identifying and defining the queries that the pilots would like to address to the Polifonia Knowledge Graph (PKN). The Survey helps to identify some of the pilot requirements and expectations. However, a more discursive and interactive approach has been designed to enter into dialogue with the pilots and to identify as closely as possible the scientific questions and the technical needs and requirements that they imply.

## 2.1 Methods and strategies

Polifonia's overall approach for centralizing needs and requirements is based on agile methodologies, eXtreme Design (XD) in particular<sup>29</sup>, and focuses on the definition of Personas, Stories and CQ. While Personas are user profiles (researcher, music lover, physically impaired person, etc.) characterized by their occupations, knowledge and interests, Stories describe the general objectives and the particular scenario of the research undertaken by this person<sup>30</sup>. The CQ are the queries that the Persona seeks to answer in the Story<sup>31</sup>.

A central concern of this methodology lies in its ability to create an interface between the disciplinary and computational dimensions of modelling: Queries are first formulated through Stories in natural language by domain experts (musicologists, historians, linguists, etc.) before being modelled by the OD team. Disciplinary technicality and computer technicality are thus distributed without being isolated. The formulation of queries in natural language by the domain experts ensures that the modelling is really driven by disciplinary problems, questions and needs. Conversely, the transposition of the queries into computer language by the ontology design (OD) team ensures that the modelling complies with current computer practices and standards. Since the approach was described in detail in D1.1<sup>32</sup>, the focus here will be on the methodological changes, the results obtained and the lessons learned since then.

To facilitate the OD activities, increase the usability of their outputs by the pilots, and articulate even more effectively the disciplinary and technological dimensions of the modelling process, the OD workflow has been extended during the last months. The new framework removes some complexity from the Story creators, while leveraging their domain expertise/experience to conveniently contextualize the specific technical scope of a Story. In particular, when defining (or refining) a Story, the creator should only focus on explaining the scope and goals of the Story itself – which task the Persona intends to achieve – and most importantly, provide practical examples and/or data to comprehensively explain the goals, needs, and artefacts of the Story. On the other hand, the Story creator is freed from defining any CQ – a responsibility which is entirely up to the OD team. Therefore, the resulting methodology is example- and/or data- driven. This has clear advantages,

---

<sup>29</sup> Kent L. Beck and Cynthia Andres. *Extreme programming explained - embrace change*, The XP series, Addison-Wesley, 2005<sup>2</sup>; Russ Unger & Carolyn Chandler. *A Project Guide to UX Design: For user experience designers in the field or in the making*, New Riders, 2012.

<sup>30</sup> The current Story template used in Polifonia is available in appendix 5.

<sup>31</sup> On the status and role of QC in the design process, see Michael Grüninger & Mark S. Fox, "The role of competency questions in enterprise engineering", in *Benchmarking - Theory and practice*, ed. by A. Rolstadås, Springer, 1995, pp. 22-31.

<sup>32</sup> See Christophe Guillotel-Nothmann *et al.* D1.1. Roadmap and pilot requirements 1st version, Polifonia, 2021, section 1.1.

as it: a) does not bias the definition of the CQ, b) promotes their consistency and reuse in other/related Stories (not necessarily from the same Persona), c) allows for the definition of objective and realistic ontological requirements, and d) makes the resulting ontologies more usable, as they are aligned *a priori* with the specific ontological needs of the pilot. On a larger scale, the consistent integration of CQ across pilots also contributes to the modular reuse of the ontology network and naturally achieves a level of interlinking among the pilot-specific modules.

To initiate the new workflow, the INTERLINK pilot has introduced a new Persona, Linka<sup>33</sup>, and a new Story<sup>34</sup> in collaboration with the OD team. As can be seen from example 4, reporting an excerpt of the new Story (*Linka#1\_MusicKnowledge*), a new section “Example Data” provides both examples of relationships pertaining to the Story (e.g. “Penny Lane is a musical work by The Beatles”) as well as relevant links to online data (e.g. MusicBrainz, Wikidata). The “Cataloguing information” section provides a number of basic assertions in natural language: “Penny Lane” is a musical work by “The Beatles”, “Penny Lane” was composed by “John Lennon” and “Paul McCartney”, etc. These assertions constitute the starting point from which more detailed discographic explorations are undertaken: The album “Strawberry Fields Forever” has been originally released in the UK on 1967-02-17, The background and lead vocals of “Penny Lane” are by “Paul McCartney”, etc. Based on the example data, the OD team produces the QCs that are the point of departure for the actual modelling. The modelling itself is then submitted to a testing tool and protocol that identifies inconsistencies and ensures that the formalization actually meets the requirements<sup>35</sup>.

---

<sup>33</sup> <https://github.com/polifonia-project/stories/tree/main/Linka:%20Computer%20Scientist>.

<sup>34</sup> [https://github.com/polifonia-project/stories/blob/main/Linka:%20Computer%20Scientist/Linka%201\\_MusicKnowledge.md](https://github.com/polifonia-project/stories/blob/main/Linka:%20Computer%20Scientist/Linka%201_MusicKnowledge.md).

<sup>35</sup> The protocol and testing tool are extensively described in Meroño-Peñuela, Albert *et al.* D.2.7. *Ontology testing and evaluation report; knowledge graph documentation and access interfaces – 1st version*, Polifonia, 2022.



## Example Data

For simplicity, music-related data can be categorised as: cataloguing information, musical features, musical facts, and musical relationships.

### Cataloguing information

Information needed to identify a musical work as well as its relationships with the recordings that are derived from it. For example:

- "Penny Lane" is a musical work by "The Beatles",
- "Penny Lane" was composed by "John Lennon" and "Paul McCartney";
- The first recording of "Penny Lane" was released in February 1967 as a double A-side single with "Strawberry Fields Forever".

Linka is interested in tracing the basic discographic information associated to the recordings of a certain musical work, including: the year of release, the name of the release, the place of recording, the instruments/performers relationships, and the vocals. Example data to address this category can be found from [MusicBrainz](#) and [Wikidata](#).

- The specific track of "Penny Lane" featured in the album "Strawberry Fields Forever" was recorded in "Abbey Road Studios: Studio 2" in "St John's Wood, Westminster, London (UK)" on 1966-12-29, 1967-01-05, 1967-01-06, 1967-01-09, and 1967-01-17.
- The album "Strawberry Fields Forever" has been originally released in the UK on 1967-02-17.
- The background and lead vocals of "Penny Lane" are by "Paul McCartney".
- The "tambourine" is played by "Ringo Starr".

### Example 4: Excerpt of the Story Linka#1\_MusicKnowledge.

The involvement of external stakeholders in the design of the pilots is promoted at two levels – a) the Survey, b) the Validation framework – through the tools set up in the Roadmap.

a) The Survey has put in place questions that aim not only to identify the communities that are relevant to the pilot, but also the (potential) partner institutions and their involvement, the way user feedback is considered, and how users are to be recruited/solicited. The questions clarify the relationship between the pilots and the stakeholder networks, address the nature of the public demonstrators that may be made available at an early stage, and ask the pilots to identify stakeholders from different institutional backgrounds and countries<sup>36</sup>.

b) The Validation framework that the pilots will apply for the mid-term (M 24) and final (M 40) validation places user feedback at its centre. Evaluation by cultural heritage communities – both amateurs and experts (researchers, teachers, students, curators, interpreters) – plays a key role for almost all pilots for the validation of a) the tools and means implemented and b) the pilot's objectives and results (see below section 4.2). The Validation framework is thus a key tool for collecting feedback from stakeholders and actively involving them in the design of the pilot and the assessment of the results produced.

## 2.2 Results

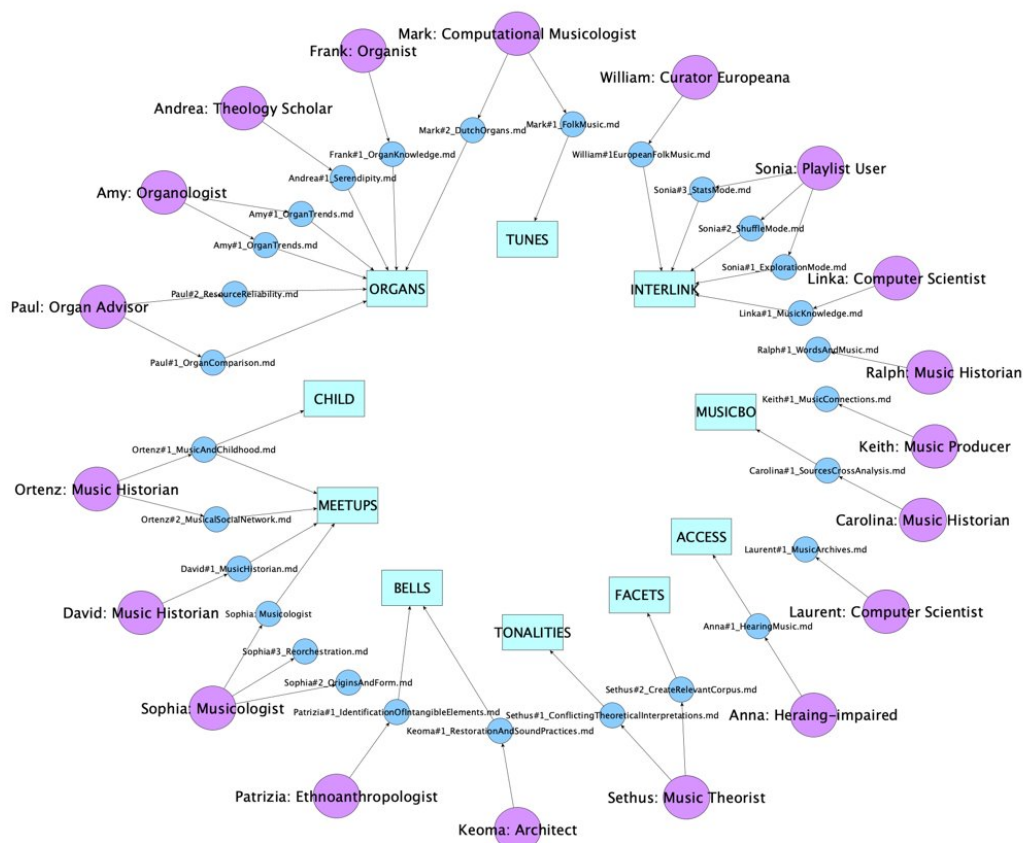
In all, the Polifonia pilots have produced 19 Personas, 28 Stories, and more than 190 CQ. Example 5 makes explicit the links that exist between a) the pilots, b) the Stories and c) the Personas. The Personas with their related Stories focus on varied problems that the pilots aim at addressing such as organ building (Amy#2\_OrganBuilders<sup>37</sup>), experiencing live music with the help

<sup>36</sup> See also above section 1.

<sup>37</sup> [https://github.com/polifonia-project/stories/blob/main/Amy:%20Organologist/Amy%232\\_OrganBuilders.md](https://github.com/polifonia-project/stories/blob/main/Amy:%20Organologist/Amy%232_OrganBuilders.md).

of haptic technology (Anna#1\_HearingMusic<sup>38</sup>), finding information about a composer related to the city of Bologna (Carolina#1\_SourcesCrossAnalysis<sup>39</sup>), the socio-cultural context of musical performances (David#1\_MusicHistorian<sup>40</sup>), compiling music programs for festivals (Keith#1\_MusicConnections), the relationship of Dutch popular music to French opera (Mark#1\_FolkMusic<sup>41</sup>), etc.

While the pilot with its particular scientific problem is the main entry point for the creation of a Persona with its specific Story and QC, it is common for these to acquire a higher level of generality later on. Thus, the Persona of Sethus, Music Theorist is used by both Tonalties and Facets, the Story Ortenz#1\_MusicAndChildhood is used by both Meetups and Child, etc. This pooling is becoming even more systematic at the level of individual QC. From a methodological point of view, the approach thus allows for the sharing and reuse of requirements.



### Example 5: Polifonia's pilots in relation to the Persona and Stories.

<sup>38</sup> [https://github.com/polifonia-project/stories/blob/main/Anna:%20Hearing-impaired/Anna%231\\_HearingMusic.md](https://github.com/polifonia-project/stories/blob/main/Anna:%20Hearing-impaired/Anna%231_HearingMusic.md).

<sup>39</sup> [https://github.com/polifonia-project/stories/blob/main/Carolina:%20Music%20Historian/Carolina%231\\_SourcesCrossAnalysis.md](https://github.com/polifonia-project/stories/blob/main/Carolina:%20Music%20Historian/Carolina%231_SourcesCrossAnalysis.md).

<sup>40</sup> [https://github.com/polifonia-project/stories/blob/main/David:%20Music%20Historian/David%231\\_MusicHistorian.md](https://github.com/polifonia-project/stories/blob/main/David:%20Music%20Historian/David%231_MusicHistorian.md).

<sup>41</sup> [https://github.com/polifonia-project/stories/blob/main/Mark:%20Computational%20Musicologist/Mark%231\\_FolkMusic.md](https://github.com/polifonia-project/stories/blob/main/Mark:%20Computational%20Musicologist/Mark%231_FolkMusic.md).

Besides their utility per se, the new Persona and Story introduced by the INTERLINK pilot were intended to provide an example following the new data-/example- driven methodology. All the other pilots were thus invited to transition accordingly, by revising their current Stories based on this example. This was also done during the WP2 Knowledge graph pilot clinics – a forum organized by WP2 where the OD team interacts with a specific pilot every week (see section 3, below).

During the initial phase of the project, the collection and updating of the Stories and their requirements was done asynchronously and according to the rhythm of each individual pilot. As the project progressed, a number of guiding elements were introduced to regulate the pace of requirements updating, such as the Maninpasta, pilot clinics and weekly cross-WP working groups (see section 3 below).

It is worth remarking that the use case of Linka already intersects common ontological needs from all the other pilots, and this has currently brought the OD team to focus on such requirements. However, out of all 19 Personas, Linka is still the only one that fully aligns with the new methodology. This suggests that the transition remains difficult, as the other pilots are still working on addressing this change.

To make the transition more fluid, a number of measures have been adopted in recent weeks. In particular, a Maninpasta organized on 16 June 2022 has contributed to refine the new workflow and to specify its four objectives: 1. making sure that ontologies check all the quality boxes, 2. making sure that pilots follow the XD methodology, 3. ensuring that pilot ontologies are integrated within PON, and 4. improving efficiency in the OD. According to these objectives, the foundations have been laid for the publication of an OD Rulebook that will provide guidelines for implementing the workflow and publishing ontologies. It seems reasonable to expect that the OD Rulebook together with the ongoing weekly working sessions (see below section 3) will foster the transition and lead to the intermediate (M24) and final (M40) validation of the pilots, which will definitively endorse the collection and validation of requirements.

Regarding the centralization of external needs and expectations, the pilots have clarified their strategies for involving external collaborators through their responses to the Survey (see table 6). This involvement is done through the validation of user scenarios (Tonalities, Child, Meetups), interfaces and/or applications (Tonalities, Child, Meetups, Bells, Interlink, Facets, Organs and Tunes) or the results produced (Interlink, Facets). It also becomes clear from the answers that the pilots have the ambition to have the Stories validated by experts (Tonalities, Child Meetups). Finally, some pilots wish to involve experts in the creation of controlled vocabularies (Bells), the identification of musical research patterns (Facets) or the identification of analytical criteria to take into account (Tonalities). To collect user feedback, adapt the tools and bring them in line with the real needs of the targeted communities, the pilots were encouraged to make demonstrators available at an early stage. The pilots thus plan beta versions of applications (Musicbo), interfaces with reduced functionalities (Tonalities, Meetups, Bells, Organs), dashboards (Organs and Tunes), case studies (Interlink) or docker versions of the tools created (Facets). These demonstrators will be made available in the coming months, some of them before summer 2022, the other during the 5<sup>th</sup> project meeting in September in Paris or during the European Research Night in September.

	Beyond the lead institution, who are the partners of the pilot and what is their involvement?		How is user/expert feedback to be taken into account in the pilot?		What kind of public demonstrators of the pilot could be made available in the coming months?
MUSICBO		MUSICBO	We plan a validation of the QC and the mock-up by external experts	MUSICBO	We plan a first demonstration of the Pilot - through a beta version of the app - during the European researchers' night in September
TONALITIES	CNAM: NEUMA score library made available and adapted to include copyright/licence information. KCL: Providing support for the modelling of theories. OU: Draft collaborative annotation interface mock-up; Identification of possible representation types/modes to be adopted; Critical discussions on theoretical models.	TONALITIES	We plan a validation of the QC and the mock-up by external experts, specialists in music analysis and theory. We actively involve experts in the identification of analytical patterns/criteria to be implemented	TONALITIES	An alpha version of the interface with limited annotation features.
CHILD		CHILD	We plan a validation of the QC and the mock-up by external experts	CHILD	None, the UI is the Listening Experience Database
MEETUPS		MEETUPS	We plan a validation of the QC and the mock-up by external experts	MEETUPS	An alpha version of the interface with limited features.
ACCESS		ACCESS		ACCESS	
BELLS	*local superintendences of Ministry of Culture (Liguria, Emilia Romagna) *Association of bells ringers (Associazione Campanari Liguri, Genova Carillons). *Other agreements are under preparation with: other local superintendences (Molise) foundries (Pontificia Fonderia Marinelli); association of experts in campanology (Associazione Italiana di Campanologia)	BELLS	We are working on vocabularies with experts and practitioners; we plan a validation of the pilot (vocabularies, mockup, application) by experts and practitioners.	BELLS	A limited number of complete examples (relation between bell towers/bells/sound recordings/performing practices) through the production of catalogue sheets: A first version of the interface and with limited data extracted.
INTERLINK	*NISV at the moment, and eventually CNAM for NEUMA as well as KNAW for the TUNES collections	INTERLINK	Quantitative and qualitative evaluation is planned for assessing the quality of links established.	INTERLINK	Tracing potential source of influence of The Beatles through music similarity. Experiments in progress.
FACETS	*Iremus-CNRS (already in the Neuma platform), *NUIG *BNF. *Comment: The pilot is still in a very early development stage and the involvement of other partners is currently limited. They will be solicited in the next months.	FACETS	We plan a validation of the pilot (UI, search quality) by external experts. We aim to involve Polifonia experts (through WP3) in selecting relevant similarity measures and patterns for the search engine.	FACETS	A standalone docker version of a preliminary version of the search engine will be released before Summer 22. It will serve as the basis of future releases (e.g., with faceting).
ORGANS	*Nationaal Instituut voor de Orgelkunst (National Institute for Organ Art): Data provider. Maintainer of target database. Publisher of the Organ Encyclopaedia.	ORGANS	We are building a network of organ advisors and organists, who we will invite to evaluate interfaces. We also plan to do a broader user study, recruiting organ enthusiasts through organ new sites and organ fora and mailing lists.	ORGANS	Data extraction is mostly done. We can provide a preliminary interface to explore the extracted knowledge graph. Also a prototype of the dashboard can be presented.
TUNES	No direct partners	TUNES	We plan evaluation of the dashboard with expert users (musicologists).	TUNES	Prototype of the dashboard.

**Table 6: Survey, engagement and validation part, partner institutions, user/expert feedback and public demonstrators.**

The pilots were also encouraged to define their relationship to the stakeholder networks. Through the Survey, they were invited to explain how their work relates to the network and to add/confirm the stakeholders with whom they (plan to) collaborate (table 7). The answers given show that the pilots target collaborations with an important network of institutions (IRISA, EPFL), infrastructures (Clarin) and individual actors in the academic world. Public institutions in the field of culture (The Stables), cultural heritage institutions (BnF, Soprintendenza Archeologia Belle Arti e Paesaggio di Bologna, Museo internazionale e biblioteca della musica di Bologna, etc.) and/or institutions focusing on inclusion (Drake music) are also at the heart of the network. Finally, the pilots, and especially Interlink, cultivate fruitful relationships with industrial stakeholders (Muziekweb, Discogs, Songfacts, WhoSampled). Through the Survey (and the Validation framework, see section 4 below), the Roadmap thus contributes to coordinating and intensifying the interactions between Polifonia and the outside world.

**Table 7: Relations to the stakeholder network and stakeholder involvement in the definition/validation of the pilot**

It is not possible, at this stage, to comment on the success of the involvement of experts in the validation of the pilots. However, we can observe from the answers given that the pilots have all planned user validation, and that they have established and put into practice strategies for the recruitment and involvement of stakeholders.



## 2.3 Lessons learned

As components of the Roadmap, the Survey and the Validation framework stimulate the involvement of stakeholders and the centralization of external feedback. Independently of the finalization of the Roadmap with this deliverable, WP1 will continue, through the Survey and the fine-tuning of the Validation framework (see below section 4 and conclusion) to foster this involvement and to bring it into practice. To this end, we plan from September of this year to organize working sessions gathering the pilots and the stakeholders around questions dedicated to the pilot's review and validation.

The transition to the new methodological workflow for OD still requires efforts from all pilots to enable a more fruitful collaboration with the OD team, which, besides the advantages highlighted in Section 2.1, is fundamental to guarantee that the XD principles and methodology<sup>42</sup> relying on pattern-based design<sup>43</sup> are consistently followed. Without such level of involvement, the integration of all the pilot-specific ontologies within the PON, and the resulting creation/inter-linking of the Polifonia KG are expected to be daunting, and potentially complex. Therefore, additional efforts will be made to implement the desired transition through additional Maninpasta sessions, where pilots sharing similar ontological needs (which were identified after the first series of the WP2 pilot KG clinics) will be invited in a dedicated forum to advance the status of their Stories. These sessions will also be an opportunity to specify common guidelines for updating ontologies and pushing them towards the expected standards.

## 3. Mapping objectives and challenges from pilot specific domains to the technology provider WPs

### 3.1 Goals, methods and strategies

In Polifonia's overall architecture, as stated in the Grant Agreement, the 10 pilots are divided into four groups<sup>44</sup>. Through their associated task and deliverables, these groups constitute a driving force for the realization of four major main objectives in relation to musical heritage: 1. preserve, 2. manage, 3. study musical heritage and 4. interact with. Additionally to the work within these groups, a tool called Maninpasta (which literally means hands in the dough in Italian) was developed at an early stage of the project to move from theoretical to practical work and to foster interdisciplinary synergy during half day events. These Maninpasta events involve researchers and developers that engage in collaborative task specific activities. They are organized on a periodical basis in the manner of hackathons with the important difference that they provide a forum not only to solve technical problems but also to discuss disciplinary issues. As such the *Maninpasta* contributes to articulating the disciplinary and technical components of the modelling tasks by creating a space for interdisciplinary dialogue and creation.

---

<sup>42</sup> Eva Blomqvist *et al.* "Experimenting with eXtreme Design", in *Proceedings of EKAW 2010*, ed. by P. Cimiano, H.S. Pinto, Springer, 2010, pp. 120-134.

<sup>43</sup> Eva Blomqvist *et al.* "Engineering Ontologies with Patterns - The eXtreme Design Methodology", in *Ontology Engineering with Ontology Design Patterns*, vol. 25. *Studies on the Semantic Web*, ed. by P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, V. Presutti, IOS Press, 2016, pp. 23-50; Valentina Anita Carriero *et al.* "Pattern-based design applied to cultural heritage knowledge graphs", «Semantic Web», 12/2 (2021), pp. 313-357.

<sup>44</sup> See appendix 3.

In this decisive phase of the project, where technologies are tested, results become concrete and validation protocols are formalized, the pilots need to interact almost daily with the technology provider WP (i.e. WP 2-5). There is also a need to foster collaboration between pilots whose areas of exploration, and therefore technological needs/solutions, have affinities. This is why the 4 pilot groups and the Maninpasta have been complemented by weekly cross-pilot-WP WGs (table 8). These groups bring together pilots and different WP to promote regular technological development and methodological reflection, tailored to the pilots, and driven by their needs. The pilots focusing on music data (Tunes, Facets, Tonalities, Interlink<sup>45</sup>) interact weekly with WP3 and WP2 while the pilots exploring textual data (Musicbo, Bells, Meetups, Child) collaborate weekly with WP4 and WP2. In addition, two “generalist” groups have been created. The first, “pilot clinics”, is dedicated to the needs and challenges encountered in the creation of individual knowledge graphs. The issues raised by the pilots (2 or 3 pilots during each weekly session) can be either methodological (e.g. modelling, relevant vocabularies and ontologies, etc.) or technological (e.g. how to use tools, infrastructures, etc.). The second “generalist” group brings together WP5, WP1 and all the pilots around issues related to the Web portal. This group plays a key role in reconciling the specific objectives and challenges of individual pilots with a common presentation and exploration environment. The group thus helps to make the pilots' objectives and challenges explicit and to identify corresponding technological solutions.

Work package	Pilot(s)
WP2	All pilots (“Pilot clinics”)
WP2 + WP3	Music-related pilots
WP2 + WP4	Text-related pilots
WP5	All pilots

**Table 8: Cross pilot-wp-groups.**

The following sections show how these four groups, which constitute an important methodological innovation compared to the previous version of the Roadmap, contribute with the Maninpasta to translating objectives and challenges from pilot specific domains to the technology provider WPs.

**WP2 Knowledge graph pilot clinics.** One of our major goals is the creation of an ontology network and a knowledge graph that build on the domain knowledge and datasets coming from the pilots. An appropriate methodology for constructing knowledge graphs has been, however, subject to much study, and various different approaches exist<sup>46</sup> to create knowledge graphs from existing data. Size and complexity are key factors to consider: for small and regular KGs off-the-shelf tools

<sup>45</sup> While the Interlink pilot, focusing on music data, relates to this group, its objective of interconnecting collections grants it a more general status. For this reason, the pilot is dealt with mainly in the pilot clinics (see below).

<sup>46</sup> Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane *et al.* “Knowledge graphs.” *Synthesis Lectures on Data, Semantics, and Knowledge* 12, no. 2 (2021): 1-257.

or custom code can be sufficient,<sup>47</sup> while for large-scale (millions of items, thousands of contributors), multi-domain KGs, such as Wikidata, a distributed platform with a purposely defined user interface is necessary<sup>48</sup>. The purpose of the WP2 KG pilot clinics is to understand pilot requirements and support the decision-making processes behind tool selection and customization for the creation of musical KGs from pilot data. In these clinics, experts in the area of knowledge graph creation visit the pilots cyclically (see table 9), diving into their datasets and data challenges, and suggest methodologies, techniques, and tools to support their uplifting as knowledge graphs; as well as the relevant resources of the PON (classes, properties) to annotate them.

Date	Pilot(s)
10/02/2022	Bells
17/02/2022	Facets
24/02/2022	Organs, Tunes
10/03/2022	Tonalities
31/03/2022	Child/Meetups
28/04/2022	Tunes
05/05/2022	Interlink

**Table 9: First (partial) cycle of pilot clinics.**

**WP2+WP3+music-related pilots.** Here, our disciplinary goals include: extracting patterns from pieces of music and from corpora, distinguishing musicologically important ones, finding relationships between pieces of music based on patterns, querying/searching corpora for patterns. We have chosen to start with a bottom-up strategy. That is, for several different genres of music and different applications, multiple teams within the group (roughly corresponding to the pilots Tunes, Facets, Tonalities and Interlink) have each defined “pattern” concepts in their own way. Each team has presented preliminary work to the WP3 group for feedback, leading to common understanding of methods and challenges. Each team then implemented their own algorithms and applications (see table 10).

Date	Agenda
04/02/2022	How to organise and use this WG
11/02/2022	How to organise and use this WG List of music ontologies Create a specific plan

<sup>47</sup> Enrico Daga, Luigi Asprino, Paul Mulholland, and Aldo Gangemi. "Facade-X: an opinionated approach to SPARQL anything." *Studies on the Semantic Web* 53 (2021): 58-73.

18/02/2022	Taxonomy of musical pattern discovery Set of "pattern dimensions" WP2 Clinic and relation to this WG
08/04/2022	Tonalities requests for algorithms/expertise in machine learning
15/04/2022	Knowledge graphs, lexicons, ontologies of patterns. NUIG will present something concrete for discussion.
29/04/2022	Items from GA/SC meeting Deliverable 3.3 status update KG discussion
06/05/2022	Requirements for pattern knowledge graphs (see initial examples below). D3.3 status update ("mining patterns" versus D3.1 "pattern extraction"). Reporting items for upcoming plenary and quarterly report.
13/05/2022	Plenary reporting: D3.3 reporting Internal gathering of Pattern KG requirements (see below) How to gather further requirements for Pattern KG? Do we need a Survey of likely users, musicologists, ...?

**Table 10: WP2+WP3+music-related pilots WG agenda.**

**WP2+WP4+text-related pilots.** Our disciplinary goals include extracting information from the text-based pilots Musicbo, Meetups, Child, Bells and Organs. First of all, we have collected the pilot's corpora and have enriched them with metadata (author, title, language of the document, date of publication). We have also collected for each pilot a sample of 1000 sentences, representative of the variety of texts encountered. We then provided the corpora with the following annotations: tokenization, lemmatization, part of speech identification, entity linking, word disambiguation and annotation of musical terms. These features do not only provide us with statistics on the linguistic composition of the Corpus, but also enable an extraction of information from the pilot's text corpora at a general level. Finally, we used the pilot samples for abstract meaning representation (AMR). i.e. a semantic parsing which consists of converting a natural language text into a logical form and representing the meaning of a sentences as a semantic graph (see table 11).

Date	Agenda	Date	Agenda
20/01/2022	Check/Complete/Comment the WP4 annual report	10/03/2022	Organization of the presentation for the first Polifonia review meeting
27/01/2022	What provenance information do we want/need to extract? Facts from the pilots: identify a possible list of types of facts, evaluate feasibility, identify overlaps. T4.2 Artifact extraction: is there any useful resource from the corpus we can use?	24/03/2022	ORGANS: updates Semantic Parsing updates Corpus Sampling updates



	(15') Benchmarks for WP4-related pilots (5')		
02/02/2022	Present the work on the annotation of the corpus (10' + 5' discussion) Present the work on selecting listening experience (10' + 5' discussion) Examples from the pilots (10' + 5' discussion) Discussion: how to combine tools and requirements (15')	31/03/2022	AMR validation: updates Corpus Sampling updates Knowledge extraction Deliverables
10/02/2022	Present AMR annotations Discuss examples and CQ from the pilots for knowledge extraction supported by the annotation provided Organs: status of play and link to bells (just to make sure this work links to WP2)	06/04/2022	Deliverables Periodicals corpus: cleaning, sentence splitting, AMR annotation (Rocco) Pilots: status updates
17/02/2022	To present results of the inspection/validation of sentence-pairs and definition of validation criteria Discuss examples and CQ from the pilots for knowledge extraction supported by the annotation provided Discuss clustering approaches to select sentences relevant to the Pilots Organs: status of play and link to bells (just to make sure this work links to WP2)	14/04/2022	D4.2 corpus organization D4.2 metadata D4.2 annotation – data format WP4 Knowledge Extraction Pilots clinics
24/02/2022	Discuss AMR (abstract meaning representation) validation criteria: motif and threshold Selection of features for the fuzzy clustering: CQ and concepts selection Fuzzy clustering of CQ/sentences Status of ORGANS	28/04/2022	D4.2: Pilot corpora D4.2: corpus annotation D4.3: status D4.5: AMR annotation of the Musicbo sample
03/03/2022	Definition of the contribution for the deliverables (M18) Discuss co-reference resolution results Discuss AMR validation criteria: motif and threshold	05/05/2022	D4.2: Pilot corpora D4.2: corpus annotation D4.3: status D4.5: AMR annotation of the Musicbo sample
		12/05/2022	[D4.2] Rocco: corpus annotation [D4.2] Rocco: corpus interrogation [D4.2] Arianna: status of the report [D4.2] Eleonora: status of the corpus [D1.2] Describe the work conducted with the pilots

**Table 11: WP2+WP4+text-related pilots WG agenda.**

**WP5+WP1+all pilots.** There are two overall objectives of the cross pilot-WP1-WP5 WG meetings. First, they aim to inform WP5 of the challenges within the pilots in terms of incremental interface design, implementation and evaluation<sup>49</sup> so that WP5 can provide more targeted support. Second, they are intended to raise awareness among the WP1 pilots of Human Computer Interaction (HCI) tools, techniques and methods being applied in WP5 research which may be of assistance to the pilots. The meetings are held weekly. The agenda alternates between being set by WP1 or WP5. Topics covered in the meetings have included validation, UX methods<sup>50</sup>, WP1 Web portal, the MEETUPS, MUSICBO and BELLS pilots, focus groups and WP4 interfaces (table 12).

Date	Agenda
28/01/2022	Web portal: Competitive analysis for the Web portal design, Dashboard templates and data Stories Meetups template
08/02/2022	UX Methodology Web portal: Update: templates and dashboard Sonar++
15/02/2022	Dashboard app Meetups + MUSICBO + BELLS template
22/02/2022	Validation criteria
01/03/2022	SONAR+ - WP 5 presentation
08/03/2022	Web portal and data Story (MEETUPS) prototype
22/03/2022	Presentation on generous interfaces
29/03/2022	Questionnaire with students on habits, expectations, frustrations in music websites UX
05/04/2022	D5.2 deliverable planning
12/04/2022	Web portal mockup
26/04/2022	D1.9 Deliverable, Polifonia Text corpus
10/05/2022	WP4 interfaces
17/05/2022	Deliverable review before Polifonia general meeting

**Table 12: WP1+WP5 + all pilots, WG agenda.**

<sup>49</sup> This process of incrementally developing and testing designs is described by Gaver as Research Through Design: each design object (e.g. mock-up or prototype) commits to a set of design choices. See William Gaver. "What should we expect from research through design?", in *Proceedings of the SIGCHI conference on human factors in computing systems*, ed. by J.A. Konstan, Association for Computing Machinery, 2012, pp. 937-946.

<sup>50</sup> A number of guides have been written describing the motivations and practice of UX Design including Rex Hartson and Pardha S. Pyla. *The UX book: Agile UX design for a quality user experience*, Morgan Kaufmann, 2018; Russ Unger and Carolyn Chandler. *A Project Guide to UX Design: For user experience designers in the field or in the making*, New Riders, 2012.

### 3.2 Results

As a whole, the groups significantly stimulate collaboration between the technical and disciplinary actors of the project: they help to specify methods and workflows for the implementation of disciplinary requirements (WP5+WP1+all pilots), and lead to a weekly collaboration of a) the text-related and b) the music-related pilots with the OD team to combine exploration and modelling tasks. Finally, the ontology clinics are a central space where the OD team collaborates with the pilots to produce, adapt and test ontologies against the initial requirements and objectives.

**WP2 Knowledge graph pilot clinics.** Since datasets from pilots display a great deal of heterogeneity, the results regarding methodology, techniques and tools for them to use in the construction of their knowledge graphs is varied. For example, INTERLINK relies on a specific framework to homogenize chord data into the JAMS<sup>51</sup> data structure, and then creates knowledge graphs using SPARQL templates and SPARQL Anything<sup>52</sup>. For the creation of links, string similarity algorithms (such as Levenshtein, sequence matching, etc.) are used; albeit the resolution of ambiguous composers and works remains a challenge. TONALITIES follows an ontology-based approach in which specific data-files (e.g. in MusicXML and MEI) are directly loaded as ontology instances and used for reasoning – the results can be serialised as RDF, therefore creating Knowledge Graphs. Other pilots follow a more data-centric approach in which the data is transformed from its original format directly to RDF: FACETS uses an API providing data in MEI and JSON; TUNES a database in JSON; BELLS an Excel spreadsheet; etc. For these, either custom code, JSON-LD contexts, SPARQL Anything or Open Refine<sup>53</sup> are viable options. ORGANS is primarily a textual database, and hence a specific approach will be needed. Lastly, WP1 has proposed that all these knowledge graphs shall be connected through a metadata abstraction layer, the Pilots Knowledge Graph (see above section 1.1.3), documenting their contents and providing a useful resource for the Ecosystem and the Web portal.

**WP1+WP3+music-related pilots.** The results include the following. 1. A complete tool for pattern extraction, identification of important patterns, and identification of relationships between pieces in the context of Irish traditional music. Informal evaluation suggests that it correctly discovers related pieces. 2. A music representation suitable for querying/searching large corpora, specifically the NEUMA corpus. It is based on a system of music content descriptors leading to features suitable for standard search methods. 3. A tool for measuring harmonic similarity between pieces, based on local harmonic patterns. It uses harmonic reduction, chord decomposition, and pitch simultaneities to arrive at a representation in which harmonic similarity can be measured. The result is a network of pieces (Isophonics for pop music, the JAAH jazz harmony dataset, and Schubert Winterreise for classical music) linked by similarity, including links within corpora and between corpora.

**WP1+WP4+text-related pilots.** Results include an annotated corpus following the CoNLL-U format where each word/token is reported in tab-separated columns on one line, and sentence boundaries are indicated by blank lines. This allows to automatically extrapolate the implicit morphosyntactic, semantic and domain-specific information. The corpus has been annotated

---

<sup>51</sup> <https://jams.readthedocs.io/en/stable/>.

<sup>52</sup> <https://github.com/SPARQL-Anything/sparql.anything>.

<sup>53</sup> <https://openrefine.org/>.

through tokenization, lemmatization, part-of-speech tagging, word sense disambiguation, named entity recognition, entity linking. We used EWISER for all languages of the corpus, ExTenD for English and mGENRE for languages other than English. The Musicbo pilot sample (1000 sentences) has been annotated with AMR.

**WP5+WP1+all pilots.** The meetings have assisted in promoting knowledge sharing and collaboration across WP1 and WP5. For example, the meetings directly led to WP5/WP1 collaboration on the nature of the focus groups to guide design of the dashboards and the Survey of music consumption in order to inform the design of the portal. Other 1-to-1 meetings have also been held or scheduled between WP5 and the pilots as a result of discussions in the WG.

### 3.3 Lessons learned

**WP2 Knowledge graph pilot clinics.** The most important lesson learned in the clinics is that the data Ecosystem of Polifonia's pilots is highly heterogeneous, and hence a unique, uniform approach for the construction of knowledge graphs, and their linking to the existing Ontology Network, is highly unlikely and reasonably discouraged. Disciplinary goals and challenges are varied among the pilots, and naturally are translated into different specific technical resources and tools. This variety comes in different shapes: size (30M triples in chord datasets in INTERLINK; to a few sheets in BELLS) is a critical dimension, but so are Personas and requirements to link with ontology classes and properties, and data formats (JSON, JSON-LD, MEI, MusicXML, text, Excel, JAMS). For the formats supported, SPARQL Anything and SPARQL query templates were successfully used to batch-convert millions of RDF triples (e.g. INTERLINK, ChoCo dataset). String similarity remains the baseline for basic entity linking, but more sophisticated methods for disambiguation, especially in music catalogs, are needed (e.g. INTERLINK, Discogs and GTAA from NISV). Various data-insights of the clinics, e.g. in BELLS, were critical to revise versions of the ontology network and correct the way resources were modelled or named.

**WP2+WP3+music-related pilots.** We have found that the bottom-up strategy already mentioned has allowed fast progress in parallel across the pilots, but a remaining challenge is to unify the multiple pattern technologies provided by WP3 to the pilots. One avenue for this unification is in current discussions of ontologies and knowledge graphs for representation of patterns. A second methodological challenge is in the evaluation of results, as we lack "gold standard" data sets labelled with "pattern importance" or "piece relationship", and moreover there is no musicological consensus on such issues. Some workarounds are being used, such as the cover-song detection task: a piece-relationship tool can be evaluated for its ability to detect cover songs or (in the case of oral musical heritage), tune variants. Moreover, such a tool based on pattern extraction can indirectly validate the appropriateness of that pattern definition. Research continues in this area.

**WP2+WP4+text-related pilots.** Some technical and methodological difficulties have been important to improve our approaches. The first concerns the annotation method: the annotation process is very time consuming, so we used textual samples representative of the linguistic variety of the pilots. This allowed us to make a first iteration of annotation in a shorter time however considering the completeness of the variety of the data. The second relates to the importance to align the semantic categories related to the analysis of musical discourse with AMR, so as to be able to maximize the extraction of information from the text following the research needs of the pilots.

**WP5+WP1+all pilots.** The meetings are providing an effective focal point for collaboration and are set to continue. One issue we may wish to consider in the future is to ensure that collaborations are appropriately captured elsewhere in the data Ecosystem, for example, capturing changes in how the users are understood as updates within the Polifonia Stories.

## 4. Monitoring pilot validation

The SMS tools in connection with the other methodological devices key to WP1 (Ecosystem, Web portal) and WP5 (XD-UX-design methodology), led to the implementation of an overall Validation framework. This section only describes the Validation framework as such and the methodological considerations behind it. The actual validation of the pilots through individual validation protocols based on the framework will be done in later deliverables at M. 24 (D1.4-7) and M. 40 (D1.8).

### 4.1 Goals, methods and strategies

The main objective of this part of the Roadmap is to design a global Validation framework that confirms that:

- O1: The pilot objectives reflect concrete needs expressed by the target communities and contribute to the state of the art;
- O2: the methods and approaches are sound;
- O3: the data on which the pilots are based are of sufficient quality to achieve the desired objectives;
- O4: the objectives and results expressed through the Stories and the Survey are achieved;
- O5: the resources, methods and tools provided are made explicit and integrated into the Polifonia Ecosystem, interconnected through the PON and conveyed through the Web portal<sup>54</sup>

The overall methodology was targeted towards a bottom-up approach that took the diversity of the pilots as a starting point and relied on the Survey, the cross-pilot-WP WGs and the Stories. This allowed common approaches and validation paths to emerge while respecting the specificities of the different pilots. As a complement to the SMS-tools, other canalizing elements including the Ecosystem with its guidelines, the XD-UX-design methodology defined within WP5, and the Web portal have been integrated in the framework. The strategy was thus not to impose homogeneity at the level of the validation protocols which will be applied by the pilots, but to create a Validation environment in which Polifonia's driving methodological principles converge.

The architecture of the Validation framework was built through a combination of Polifonia's general methodological guidelines and an in-depth analysis of the validation methods, protocols, success criteria and KPIs reported by the individual pilots in the Survey (see above section 1). On this basis, a task force composed of Polifonia members developed a general scheme that is flexible enough to accommodate the current and future needs of the pilots and binding enough to meet the objectives listed above.

---

<sup>54</sup> Only some of the pilots' data are conveyed through the webportal. See the Survey, technical part, questions "Is it to be released through the Polifonia Web portal or as an independent application?", "How should the pilot's resources be presented, linked, or stored on the website dedicated to the pilot?", "What type of knowledge can contribute to the Polifonia Web portal?", "What type of interaction do you expect should be supported by the Web portal?", "In which ways the Web portal shall link to the (separate) pilot demonstrator (if any)?" and "What kind of 'augmentation/enrichment' do you expect to get out of being displayed via the portal?". See also Marilena Daquino *et al.*, *D1.9. Polifonia Web portal – 1st version*, Polifonia, 2022.



The Validation framework itself is a Github repository in the Polifonia Ecosystem<sup>55</sup> and will become an Ecosystem component. The framework can thus be documented, adapted and monitored until its actual application in the dedicated deliverables at M24 and 40.

## 4.2 Results

The Validation framework is modular in nature. It includes core criteria with which all pilots must comply: a) the use of a methodological approach combining XD and UX design, b) the inclusion of resources in an Ecosystem governed by specific guidelines, c) the presentation of the pilots and (part of) their resources via the Web portal; and d) the regular monitoring via the Survey. In addition to these criteria, the framework consists of specific validation methods for individual pilots or pilot groups (table 13).

Core validation criteria	XD and UX Design		<div>✓</div>											
	Ecosystem + Rulebook													
	Web Portal													
	Survey (with success criteria, KPIs and validation setup)													
Pilot specific methods	Input validation	Validation of mechanically produced input data by ground truth						✓		✓				
		Quantitative validation of input data		✓					✓					
		Validation of input data quality by experts		✓	✓	✓		✓						
	Validation of tools/means/software	Validation of interfaces by users/experts	✓	✓	✓	✓	✓	✓		✓	✓	✓		
		Performance validation					✓							
		Validation of performance, requests/information passed between front and back end	✓			✓					✓			
	Output/goal valiation	Compliance of the data produced with standards		✓										
		Validation of results by ground truth					✓				✓	✓		
		Validation of results by experts					✓							
		Validation from a sub-corpus/case study and transfer to a larger corpus		✓					✓					
		Validation of objectives through experts	✓	✓	✓			✓			✓			
		ACCESS	BELLS	CHILD	FACETS	INTERLINK	MEETUPS	MUSICBO	ORGANS	TONALITIES	TUNES			

**Table 13: Validation framework, core validation criteria and pilot-specific methods.**

### 4.2.1 Core validation criteria

The compliance of the pilots to a set of common core tools, principles and guidelines grants a methodological and conceptual coherence to the whole while leaving enough space for individual pilot specific validation protocols.

**a) XD and UX design:** The strategy consists of applying a mix of eXtreme Design and UX design through user interviews, Personas, Stories, CQ, mock-ups, and interface prototypes. This approach, which is part of design-centered research (RTD) helps unify the design process of the pilots and ensures that their objectives are in line with real needs (**O1**). Furthermore, the inclusion of the Stories as key elements in the pilot validation process enables us to monitor the extent to which the requests and needs formulated through them are actually implemented (**O5**).

**b) Ecosystem and Rulebook:** The conformity of the pilots (and of their components) to the Ecosystem and its guidelines plays a leading role in the validation process. This criterion ensures that all components that are discussed, presented and published are correctly referenced (**O5**). The

<sup>55</sup> [https://github.com/polifonia-project/socio-technical\\_roadmap/tree/main/validation\\_framework](https://github.com/polifonia-project/socio-technical_roadmap/tree/main/validation_framework).

Ecosystem also makes sure that the decisions, modifications and progress related to the pilots are documented. This provides central perspectives for the maintenance of the pilots and for assessing that the Stories are considered in the pilot implementation and validation process. Finally, this validation criterion helps to interconnect the pilot components and to provide them with metadata (type, WP, pilot, keywords).

**c) Web portal:** As a place for disseminating, enhancing and interconnecting the pilots, the Web portal helps to harmonize the way Polifonia's resources can be presented, explored and searched by both amateur and expert users. It also favors co-creation processes and the mutual adjustment of the individual pilots. Methodologically, the portal helps to steer the pilots towards a story-telling, data-, and hypothesis-driven approach. The design principles based on QC analysis, focused groups to collect users' expectations and frustrations, and prototyping (based on the information collected and feedback from Polifonia researchers) lead to identify and apply shared presentation and interaction patterns. Thus, the Web portal is an important validation criterium because it contributes to applying sound methods and approaches **(O2)** and to conveying Polifonia's resources to a dedicated and shared place **(O5)**.

**d) Survey:** The key position of the Survey is due to its ability to both reveal the DNA of the pilots and to make conclusions on the architecture of the Validation framework. This is because the Survey provides meticulous documentation on the choices, orientations, resources and objectives of each pilot: musical heritage collections explored, copyright and licensing issues, technical specificities, application to be developed, relationship to the Web portal, involvement of experts, consideration of user feedback, collaborations etc. The Survey is also the main source of information for the pilot's individual validation strategies. It is the place where the monitoring of the pilots has been carried out so far and where detailed information on the validation methods, the validation protocol and the KPIs used is collected<sup>56</sup>. The Survey thus has a double role. On the one hand, it is a central tool for designing the Validation framework presented here. On the other hand, it provides us with a fingerprint of the pilots scientific, technical and socio-pedagogical specificities. This fingerprint can be used not only internally (to specify and analyze individual approaches) but also externally (for example on the Web portal, on the pilots' individual applications, or in publications). The Survey thus gives a view of the level of implication of the scientific communities **(O1)**. It also allows us to reconstruct the pilot's methodological principles **(O2)**, and to monitor the extent to which the objectives are achieved **(O4)**.

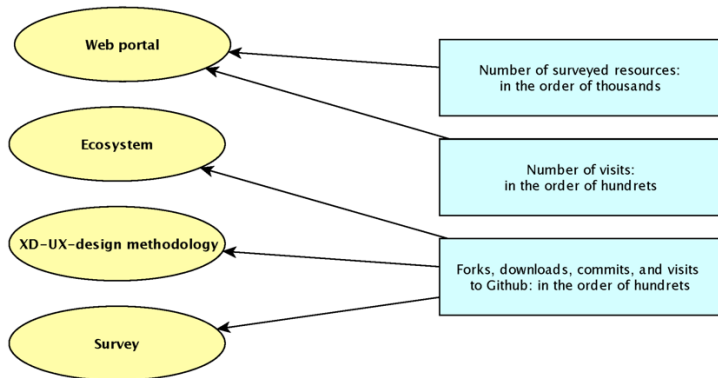
Although only the results of the pilots and of the overall project can demonstrate the effectiveness of the core tools implemented (XD and UX-design, Ecosystem and Rulebook, Web portal and Survey), these tools themselves should be evaluated and monitored, particularly with a view to adapting and transferring them to other projects. The use of user Surveys as a qualitative approach to validate these components is methodologically difficult to implement as it cannot be entrusted to external experts. It is planned instead to methodologically focus the validation of the framework's components on KPIs according to the metrics outlined in example 6. As regards the XD-UX-design methodology our strategy consists not only of demonstrating what the methodologies produce (i.e. number of components, Stories, Personas, competency questions, see above section 2.2) but also to monitor the success of the tools that have been designed on the basis of this methodology

---

<sup>56</sup> See section 1.1 above.



through the number of people testing, using and adopting them. The indicators of success currently used, which stem from the Grant Agreement<sup>57</sup>, may be extended to include additional criteria, such as the number of stars in repositories, number of external developers and peer-reviewed publications.



#### Example 6: KPI for the Web portal, Ecosystem, XD-UX-design methodology and Survey

#### 4.2.2. Pilot specific methods

The pilot specific methods identified are the result of an in-depth analysis of the answers given by the pilots during several iterations of the Survey<sup>58</sup>. First, the pilots freely described their validation methods. The answers were then reformulated in order to identify common methods and to group them into homogeneous categories. These categories were then submitted to the pilots for adjustment and validation. The validation methods identified on this basis cover a) the “input data” of the pilots, i.e. the data that serve as starting points for the case studies; b) the tools, means and software bricks by which new knowledge and/or results are produced; and c) the validation of objectives and results (see table 13).

The following paragraphs describe these validation methods in general and illustrate them with individual examples. A systematic and detailed description of the methods applied to each pilot is provided in the deliverables for the mid-term and final evaluation.

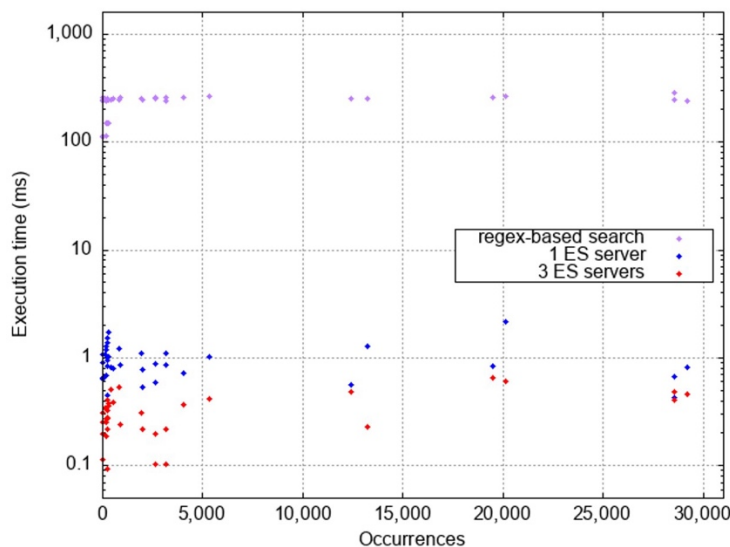
**a) Input validation.** The goal of this first validation group is to assess the quality of the data (musical scores, historical texts) with regard to pilot specific needs (O3). Another objective is to leverage the quality control of the input data to identify what changes have been made to the dataset to suit the pilot’s requirements. Validation on the basis of ground truth of mechanically produced input data concerns the scanning quality of printed documents (Organs) and the harvesting of information collected on the Internet (Meetups). Organs for example will compare the automatic conversion with manual input to provide a coefficient of correctly converted sources. In addition to this approach, two pilots (Bells and Musicbo) explicitly consider the quantity of input data as a validation criterion. For instance, Bells will collect at least 100 complete case studied and will digitize at least 300 records. The criterion most frequently used by the pilots is the qualitative validation of data by

<sup>57</sup> The KPIs currently considered are based on the Grant Agreement, part B, p. 11, table 1.1.

<sup>58</sup> <https://github.com/polifonia-project/Survey>. Iterations took place at the following dates 07.05.2021, 14.05.2021, 17.06.2021 and 19.05.2022.

experts, for example the information provided by bell ringers (Bells) or the quality of the digital scores explored (Facets).

**b) Validation of tools/means/software.** The objective here is to verify both the robustness and the ergonomics of the technical solutions implemented (**O2**). It is highly significant that the validation of the tools/media/software bricks focuses on the validation of interfaces by amateur and/or expert users. The almost complete consensus on this criterion reflects the pilots' adherence to the XD-UX-design methodology used as a core criterion. Tonalities for instance plans to run a usability survey complemented by interviews for the interface and will consider the resource to be successful when the user survey will achieve 90% positive evaluation. Facets on the other hand will take into account the number of partners or external institution/people using its tool (or collaborating to it) by measuring Github forks (collaboration), Docker image downloads, and the number of verified installations. Apart from this method, the pilots focus on the accuracy and reliability of the requests passed between the front and back end (Access, Facets, Tonalities) and on the performance of the computer solutions to be developed (Facets). To this end, Facets measures the accuracy and the average time the search engine (i.e. elastic search) needs to answer a query compared to other search strategies (example 7).



**Example 7: Comparison of execution times of regex-based searches and elastic search (ES) servers**

**c) Output/result validation.** The last subgroup covers the validation of both the objectives and the results of the pilots (**O4**). One part of the pilots validates its results based on the ground truth previously created. This validation method concerns the music-related pilots Interlink, Tonalities and Tunes. Tonalities for example aims at providing 90% correct modal-tonal identifications in a corpus with at least 150 works with one analysis. The strategy adopted in the case of Bells and Musicbo, both focusing on text, is to validate a larger set of results from narrow case studies: information gathered about a small set of bells or a specific aspect of Bologna's musical heritage. Two pilots (Bells and Interlink) consider respectively the conformity of the data to standards and the validation by experts of the results obtained (i.e. the interconnections made between different musical works). Interlink for instance plans to rank the similarities detected between musical pieces based on known categories (performances of the same composition, covers songs, etc.) according to f-score and to

compare these results with human performances. The validation of the objectives by experts is, however, here again, the most important criterion: Access for instance considers the views of music educators and music venues involved while Child takes into account the declared interest of the Music Department in Open University in tools and methods defined.

#### 4.3 Lessons learned

An important effort has been made to build a coherent validation infrastructure. This effort is both individual and collective: The pilots have defined on their own their validation strategies according to what is best suited to their needs. At the same time, the cross-pilot-WP WGs and the task force activities related to the pilot validation have contributed to designing an overall Validation framework combining core validation criteria and pilot specific methods organized into three subgroups.

In all subgroups, validation by users is a central criterion. This general focus is in line with and fosters the project's overall commitment to obtaining early feedback from musical heritage communities, to avoid technology driven development and to foster user-centered approaches. By doing so, we intend to create a virtuous circle between the users' feedback and the subsequent adaptation/refocusing of the pilots. The validation process has a central role to play here since its application will lead to a substantial increase of the feedback obtained. At the same time, the information collected from the Survey suggests that there is a need to map user categories more systematically to the pilot's general goals and, more specifically, to the validation tasks associated with them: preserve (Bells, Organs), manage (Facets, Interlink), study music heritage (Tonalities, Tunes, Child, Musicbo), and interact with it (Meetups and Access).

The methods that relate to the validation of outputs and results identify some affinities between Interlink, Tonalities and Tunes on the one hand and Musicbo and Bells on the other. However, it is not possible at this stage to identify pilot sets that would share homogeneous validation methods (e.g., text-related pilots and music-related pilots). This may be because the responses are not yet systematic enough and/or because the pilots, depending on their scientific perspectives and objectives, grant unequal importance to the different validation methods. At least one new iteration of the Survey before the mid-term validation (M24) is scheduled to update the information. Furthermore, it is planned to integrate the different WGs more actively in the process of identifying shared methods in the coming months.

Finally, it appears necessary to generalize the interconnections between the CQ and their effective integration and implementation in the pilots. The bi-directional link between CQ and pilot updates is taken into account in the current validation infrastructure through the XD-UX-design methodology at the level of the "core criteria". However, this link still needs to be concretely implemented in the Ecosystem in the coming months before the intermediate validation.

## FAIR protocol

The following components have been released as a direct or indirect result of the work described in this deliverable:

1. Ecosystem
2. Web portal
3. XD-UX methodology related data
4. Survey
5. Validation framework

Via Github, all components – or in the case of the XD-UX-design methodology their outcomes (i.e. the Personas, Stories and CQ) – have a unique and persistent identifier and are registered in a searchable resource. The Survey, in its current version, has also been deposited in Zenodo<sup>59</sup>. All the components have been described with and/or contribute to produce rich metadata: The Ecosystem is targeted towards producing metadata through its annotation schema; all resources of the Web portal are provided with metadata (scope, audience, type of data served, data access points)<sup>60</sup>. Finally, as components of the Ecosystem, the Survey and the Validation framework are described with metadata defining their type, contents, and relation to the WPs. This also applies to the outcomes of the XD-UX-design methodology. The storage of all resources on Github and on Zenodo (in the case of the Survey) guarantees their persistence beyond the project.

Since all components and their related data benefit from the Ecosystem annotation schema (and are registered in Zenodo in the case of the Survey), they are all retrievable by their identifier using standardised communications protocols. These protocols are open, free, and universally implementable. Furthermore, the Ecosystem infrastructure with its annotation schema is independent of its individual components. Therefore, the metadata remains accessible, even if the data is no longer available.

All data and metadata produced use a formal, accessible, shared, and broadly applicable language for knowledge representation. The resources disseminated by the Web portal are made available in the form of RDF triples from the PKG. The Survey data will also eventually be transformed into an RDF knowledge graph (the PiKG). The resources created by the XD-UX design methodology through the Persona, Stories and CQ are all open linked data of the PON. As for the Ecosystem and Validation framework, their organisation is made explicit by the Ecosystem annotation scheme.

Finally, the Ecosystem, the Survey, the Stories which are derived from the XD-UX-design methodology, and the Validation framework are released with a CC-BY 4.0 licence that allows all resources to be copied and redistributed in any medium or format. The material may also be remixed, transformed, and build upon for any purpose, even commercially. The different resources of the Web portal are released under CC0 and ISC licences.

---

<sup>59</sup> Guillotel-Nothmann, Christophe, Gurrieri, Marco, van Kranenburg, Peter, Musumeci, Elena, Veninata, Chiara, Berardinis, Jacopo de, Fournier-S'niehotta, Raphaël, Marzi, Eleonora, Barlow, Helen, Daga, Enrico, & Holland, Simon. (2022). Polifonia Survey (Version 19062022) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.6664200>.

<sup>60</sup> Furthermore, it is planned to develop a dedicated pipeline for data enrichment for case the metadata would be incomplete.

## Conclusion

This deliverable describes the work done over the last 12 months to set up the second version of a socio-technical Roadmap. This Roadmap consists of a common methodological framework on which the pilots, as the driving force behind *Polifonia*, can build. In the first version of the Roadmap, our efforts focused on coordinating the collection of the pilot's expectations and needs through a) the Stories, b) the Maninpasta, and c) the Survey (SMS) in a bottom-up approach. Based on an in-depth analysis of the information provided through the Survey and considering the methodological approaches that crystallized through the Stories and the Maninpasta/cross-pilot-WP WGs, this second version of the Roadmap presents a generalization of the methods, standards and procedures which culminates in the Validation framework. To this end, the collection of the pilot's expectations, which continued during this year, has been integrated into a broader set of objectives:

1. Coordinating resources, methods and tools used in the pilots
2. Centralizing needs and requirements of internal and external stakeholders
3. Mapping objectives and challenges from pilot specific domains to the technology provider WPs
4. Monitoring pilot validation

1. The coordination of the resources used by the pilots through a) the Ecosystem and its guidelines, b) the Web portal and c) the Survey has led to structured, homogeneous and highly granular information on the pilots' resources, methods, tools, principles and guidelines (including the Personas, Stories and CQ relating to the pilots). The tools have helped to uncover interconnections between the pilots and to identify the resulting opportunities for resource sharing. The Survey, once transformed in the PiKG has the ambition to be the focal entry point to understand the pilots in their domain-specific, technical and socio-pedagogical dimension.

2. Stories have proven to be a suitable tool for centralizing the pilot's requirements and needs. A systematic analysis of the 19 Personas, 28 Stories and 192 CQ currently available has allowed to define the PON, to design each module and its expected inter-relationships, and to implement the basic features of each module. The centralization of external feedback, especially from stakeholders, will be provided through the Validation framework, which integrates the XD-UX-design methodology as a core criterium and focuses on user feedback through the individual validation methods.

3. Maninpasta and, at a later stage, the cross-pilot-WP WGs have proven successful in mapping challenges from pilot specific domains to the technology provider WPs. They have led to common tools and methods adapted to the exploration of textual and musical data and have contributed, through the pilot clinics, to build a direct link between the pilot's specific domains of expertise and the modeling of these domains in the PON.

4. The detailed analysis of the Survey and the work carried out in the cross-pilot-WP WGs allowed for the development of a Validation framework combining core validation criteria and pilot specific validation methods. This framework is the cornerstone of the pilot-driven work conducted in WP1 in conjunction with the technology provider WPs and leads to a harmonization of methods, practices and approaches.

While the socio-technical Roadmap is completed with this deliverable, various actions directly or indirectly related to the Roadmap are planned for the coming months. We will continue to populate the Survey and eventually will transform it into a KG to make it an interconnected, browsable and explorable component for both internal and external purposes. Another field of activity – taking place in WP2 but directly connected to the objective of centralizing needs and requirements – consists of reformulating the Stories and carrying out their systematic analysis. We will actively engage the network of stakeholders in the validation process by collecting their feedback through user studies, Surveys, reports and dedicated WGs. Finally, in conjunction with the pilots and the cross-pilot-WP WGs, the fine-tuning of the Validation framework will be pursued to apply it effectively in the intermediate (M24) and final (M40) pilot evaluations.



## References

- Beck, Kent L., & Andres, Cynthia. *Extreme programming explained - embrace change*, Addison-Wesley, 2005<sup>2</sup>.
- Blomqvist, Eva *et al.* "Experimenting with eXtreme Design", in *Proceedings of EKAW 2010*, ed. by P. Cimiano, H.S. Pinto, Springer, 2010, pp. 120-134.
- Blomqvist, Eva *et al.* "Engineering Ontologies with Patterns - The eXtreme Design Methodology", in *Ontology Engineering with Ontology Design Patterns*, vol. 25. *Studies on the Semantic Web*, ed. by P. Hitzler, A. Gangemi, K. Janowicz, A. Krisnadhi, V. Presutti, IOS Press, 2016, pp. 23-50.
- Carriero, Valentina Anita *et al.* "Pattern-based design applied to cultural heritage knowledge graphs", «Semantic Web», 12/2 (2021), pp. 313-357.
- Daga, Enrico *et al.* *D1.3. Pilots development – collaborative methodology and tools*, Polifonia, 2021.
- Daga, Enrico, Luigi Asprino, Paul Mulholland, and Aldo Gangemi. "Facade-X: an opinionated approach to SPARQL anything." *Studies on the Semantic Web* 53 (2021): 58-73.
- Daquino, Marilena *et al.*, *D1.9. Polifonia Web portal – 1st version*, Polifonia, 2022.
- Gaver, William. "What should we expect from research through design?", in *Proceedings of the SIGCHI conference on human factors in computing systems*, ed. by J.A. Konstan, Association for Computing Machinery, 2012, pp. 937-946.
- Grüninger, Michael & Fox, Mark S., "The role of competency questions in enterprise engineering", in *Benchmarking - Theory and practice*, ed. by A. Rolstadås, Springer, 1995, pp. 22-31.
- Guillotel-Nothmann, Christophe *et al.* *D1.1. Roadmap and pilot requirements 1st version*, Polifonia, 2021.
- Hartson, Rex, & Pyla, Pardha S. *The UX book: Agile UX design for a quality user experience*, Morgan Kaufmann, 2018.
- Hogan, Aidan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane *et al.* "Knowledge graphs." *Synthesis Lectures on Data, Semantics, and Knowledge* 12, no. 2 (2021): 1-257.
- Meroño-Peñuela, Albert *et al.* *D.2.1. Ontologies and knowledge graphs of music objects, patterns, and software package – 1st version*, Polifonia 2021.
- Meroño-Peñuela, Albert *et al.* *D.2.4. Methods for interlinking knowledge graphs – 1st version*, Polifonia 2022.
- Meroño-Peñuela, Albert *et al.* *D.2.7. Ontology testing and evaluation report; knowledge graph documentation and access interfaces – 1st version*, Polifonia, 2022.
- Scharnhorst, Andrea *et al.* *D7.1. First Data Management Plan*, Polifonia, 2021.

Scharnhorst, Andrea et al. *D7.4. DMP webinar on dynamic data management*, Polifonia, 2021.

Unger, Russ, & Chandler, Carolyn. *A Project Guide to UX Design: For user experience designers in the field or in the making*, New Riders, 2012.

Vrandečić, Denny, and Markus Krötzsch. "Wikidata: a free collaborative knowledgebase." *Communications of the ACM* 57, no. 10 (2014): 78-85.

## Appendices

### Appendix 1: Glossary

Term	Definition/description
<b>AMR</b>	<b>Abstract Meaning Representation</b>
<a href="#">ArCo</a>	<b>Architettura della Conoscenza.</b> A project for developing the Italian cultural heritage knowledge graph.
<a href="#">CIDOC CRM</a>	CIDOC Conceptual Reference Model (CRM) is a theoretical and practical tool for information integration in the field of cultural heritage.
CoNLL	An annotation format, see <a href="https://universaldependencies.org/format.html">https://universaldependencies.org/format.html</a>
<b>CQ</b>	Competency Question: a question that an ontology would need to answer.
<a href="#">CRM<sub>inf</sub></a>	CRM <sub>inf</sub> : the Argumentation Model. An Extension of CIDOC-CRM to support argumentation.
<b>D</b>	Deliverable.
<a href="#">Discord</a>	Instant messaging and digital distribution platform designed for creating communities.
<b>DMP</b>	Data Management Plan.
<b>Domain specific part/aspect</b>	Any scientific, epistemological and methodological considerations concerning one or more pilots.
<b>EWISER</b>	Enhanced WSD Integrating Synset Embeddings and Relations.
<a href="#">ExTenD</a>	Extractive Entity Disambiguation.
<b>eXtreme Design (XD)</b>	Pattern-based ontology engineering methodology, which collects requirements in the form of Competency Questions.
<a href="#">Github</a>	Git is an open-source software for managing versions of documents. The documents typically managed in Git repositories are computer source code files, but it is in fact perfectly suitable for text data in general, including humanities data. The development of Git is closely related to that of <a href="#">GitHub</a> a service for hosting Git repositories.
<b>Goal</b>	A short textual description of the goal(s) that the Persona needs to be addressed in the Story.
<b>H</b>	Human(s).

<b>HCI</b>	Human Computer Interaction.
<a href="#">IReMus</a>	Institut de Recherche en Musicologie, UMR 8223. Joint research unit composed of the CNRS, the BnF, Sorbonne University and the French Ministry of Higher Education and Research.
<b>KG</b>	Knowledge graph.
<b>Knowledge unit/type</b>	New knowledge produced by the pilot through the analysis (by algorithms, machine learning and/or human annotation) of the input data. This unit may correspond for example to a cadence in a score, to the identification of a previously anonymous author of a source, to the type of a bell, etc. (see diagram 2. Survey – Knowledge units investigated by the pilots).
<b>mGENRE</b>	A sequence-to- sequence system for the Multilingual Entity Linking.
<b>M[1-40]</b>	Month counting from the start of <i>Polifonia</i> on 1 January 2021: M6 (= June 2021), M18 (= June 2022), etc.
<a href="#">MEI</a>	Music Encoding Initiative.
<b>MH</b>	Musical heritage.
<b>NLP</b>	Natural language processing.
<b>Observation unit</b>	A portion of input data on which the pilot relies to provide new units of knowledge through algorithms, machine learning and/or human annotation. These observation units may correspond for example to a sentence in a text, a score fragment (for example a cadence) or the dedicatee engraved on a bell.
<b>OD</b>	Ontology design.
<b>ODP</b>	Ontology design patterns which correspond to small ontologies that work as reusable solutions to recurrent modelling problems.
<b>OEI</b>	Office of ethics and integrity.
<a href="#">OU</a>	The Open University.
<b>OWL</b>	Web Ontology Language (OWL) is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things.
<b>Persona</b>	A research-based description of a typical user. This description contains attributes such as name, age, occupation, and relevant characteristics such as knowledge and skills.

<b>PIKG</b>	Pilot Knowledge Graph.
<b>PKG</b>	Polifonnia Knowledge Graph.
<b>PON</b>	Polifonia Ontology Network.
<b>POPD</b>	Protection of personal data.
<b>RTD</b>	Research through design.
<b>Scenario</b>	A description of how in a Story the Persona's task/need/problem is solved before, during and after interaction with the resource/software/service being developed.
<b>SMS tools</b>	The three tools at the core of this deliverable: Story, <i>Maninpasta</i> and Survey.
<b>Socio-pedagogical aspect</b>	An item linked to a holistic and relationship-centred way of working in care and educational settings with the possible final users of <i>Polifonia</i> .
<b>Source</b>	Within a dataset, it is meant to define the origin of the input data, in opposition to the output data, that are the results collected by the pilot.
<b>Story</b>	A Story is a template for collecting requirements.
<b>TB</b>	The Technical Board, within the WP1, is in charge of the technical development, making sure that synergies are identified in a timely manner and issues around compatibility and technical interoperability are addressed harmonically during the development phase. The TB defines the common methodological approach and identifies the supporting collaborative tools to be used.
<b>UI</b>	User interface.
<b><a href="#">UNIBO</a></b>	Alma mater studiorum – Università di Bologna.
<b>URI</b>	A Uniform Resource Identifier is a unique sequence of characters that identifies a logical or physical resource used by web technologies.
<b>UX</b>	User Experience Design.
<b>WG</b>	Working group.
<b>WP</b>	Work package.
<b><a href="#">XD</a></b>	Extreme design, a framework for pattern based ontology design.

## Appendix 2: List of the pilots

<b>Polifonia's Pilot</b>	<b>Action</b>
<a href="#"><u>BELLS</u></a>	Bell structures are widespread both in urban and rural areas. They contribute to the distinctive shape of a landscape, to defining its soundscape and play as markers of daily, festive and ritual times. Bell heritage is complex and fascinating and influences our perception of the places we live daily. Both its tangible and intangible assets, and their dependencies are hardly encoded explicitly: most of this heritage is transmitted orally. This pilot intends to encode this valuable information in a knowledge graph, which will be publicly available and particularly relevant for scholars and cultural institutes.
<a href="#"><u>ORGANS</u></a>	The history of pipe organs is rich and diverse, and highly interrelated to economic, religious and artistic contexts. Currently, the information about building practices and characteristics of ~2000 Dutch pipe organs is only retrievable by manually paging a 15 volumes (4,500+ pages) encyclopaedia: the Orgelencyclopedia (1997-2010). This pilot will build a knowledge graph out of the text of Orgelencyclopedia, which will provide digital (and quick) access to such huge and detailed knowledge, including connection to data about aspects of their wider historic contexts.
<a href="#"><u>FACETS</u></a>	Music libraries currently lacks well-founded information retrieval tools. This pilot will design a faceted search engine (FSE) for music score collections, supporting explorations and discovery of scores of interest in large collections, based on features such as melodic, harmonic or rhythmic patterns, style, structure, instrumentation, metadata leveraged at the collection level.
<a href="#"><u>INTERLINK</u></a>	In order to answer research questions, musical heritage scholars need to combine diverse datasets (music scores, audiovisual materials, metadata) from European digital music libraries and audiovisual archives. They need to identify common entities and concepts implicitly present in the data, across different collections in different institutions. This process is mainly conducted manually in isolation and the diverse results are rarely connected. This pilot will focus on revealing and make compatible the entities and concepts hidden in digital music libraries and audiovisual archives.
<a href="#"><u>CHILD</u></a>	This pilot will build a knowledge graph of the historical experience of music in childhood, using life writing (letters, diaries, memoirs, travel writing) and other historical texts as sources for adult reflections on music heard in childhood, third-party observations on children's engagement with music, and children's own first-hand accounts. The resulting knowledge graphs will inform an interface enabling the exploration and analysis through the dimensions of themes, time, and space.



<a href="#">MUSICBO</a>	Music has always played a central role in the city of Bologna. Nevertheless, its musical heritage is only partly known and enjoyed compared to its full potential. This pilot will create multilingual (English, French, Spanish and German) digital corpora, containing the testimonies of scholars, journalists, travellers, writers and students from medieval to modern times through published documents showing diverse discourse styles such as Stories, letters, reports, news, reportage, etc. These corpora will be the base to build a knowledge graph available to researchers, cultural institutes and public administrations for reuse.
<a href="#">TUNES</a>	The digital music collection of the Meertens Instituut (Amsterdam) includes thousands of melodies from Dutch popular culture, spanning a period of more than five centuries. To trace possible international origins of Dutch early popular music culture, this pilot will interlink the entire melody collection of the Meertens Institute with a large number of other European collections. The linked melodic data sets will be highly valuable for musicologists and music historians interested in cultural evolution of musical style and in oral transmission and variation.
<a href="#">TONALITIES</a>	The modal-tonal organisation of Western music is decisive for its inner coherence, its dramatic plot and, ultimately, for its artistic meaning. This pilot develops tools for the modal-tonal identification, exploration and classification of monophonic and polyphonic notated music from the Renaissance to the 20th century.
<a href="#">ACCESS</a>	The aim of this use case is to co-design, develop and evaluate wearable haptic technology to enable people who are Deaf or hearing impaired to engage as audience members in live performances.
<a href="#">MEETUPS</a>	This pilot focuses on supporting music historians and teachers by providing a Web tool that enables the exploration and visualisation of encounters between people in the musical world in Europe from c.1800 to c.1945, relying on information extracted from public domain books such as biographies, memoirs and travel writing, and open-access databases. These encounters will be explored in a timeline and map interface and may reveal unexpected connections and relationships that cast new light on aspects of European music history. The tool will provide persistent, citable identifiers in order to support referencing in scholarship outputs.

## Appendix 3: WP1 participants, efforts, objectives, tasks and deliverables

Title	Web portal and pilots									
Participant	1	2	3	4	5	6	7	8	9	10
	BO	OU	KCL	NUIG	ICCD	ICBSA	IREMUS	CNAM	NISV	KNAW
Person months	8	10	10	6	6	5	12	11	4	11
Start month	1			End month			40			

### Objectives

WP1 is *Polifonia*'s validation WP, it contributes to achieve O3 (Tailor), and O5 (Share & Engage). As such it drives the whole development. It also delivers a registry of all resources and materials retrieved, used and produced in the project, in the form of a Web portal. Its goal is twofold: 1) to demonstrate that the methods and tools developed in the technology provider work packages (WP2-5) are effective in facilitating management of large musical heritage collections and supporting enhanced understanding, preservation of, and interaction with, musical heritage, 2) to contribute to push the state of the art in relevant, though specific, musical heritage use cases. It includes seven tasks: T1 delivers a socio-technical Roadmap to all partners ensuring supervision for the co-creation process; T2 provides a Technical Board supervising the development; T3 delivers a Web portal / registry and associated services; T4-7 develop ten pilots. Each task represents a theme relevant to the scope and challenges of the call: Preserving MH (T3), Managing MH (T4), Studying and understanding MH (T5), Interacting with MH (T6). Notice that ten "external" early adopters have already expressed their interest in participating in seven pilots, as explicitly indicated in the pilot descriptions (see also attached support letters).

### Tasks

#### Task 1: Socio-Technical Roadmap (Leader: IREMUS - Participants: ALL)

This task provides a common framework to all pilots (T4-7) for: 1) coordinating collection of resources for use in the pilots, 2) gathering requirements from internal and external adopters/stakeholders, 3) setting objectives and challenges and mapping them from specific domains to the technology provider work packages (WP2-5), 4) monitoring validation within the pilot tasks. The goal is to maximise synergy and methodological soundness.

#### Task 2: Technical Coordination (Leader: OU - Participants: ALL)

This task will provide the pilots and the technology provider WPs with a reference Technical Board (TB) which, according to the Roadmap defined in T1, will supervise the technical development making sure that synergies are identified in a timely manner and issues around compatibility and technical interoperability are addressed harmonically during the development phase. The TB will define the common methodological approach and identify the supporting collaborative tools to be used. It will interact with WP6 to make sure that all the developed components are released

according to the FAIR protocols defined in T7 of WP6. This task will also collect and report validation results from the pilots.

**Task 3: *Polifonia* Web portal: an aggregator of digital musical heritage collections** (Leader: UNIBO - Participants: OU, NUIG, KCL)

By extending musoW, a catalogue of Musical Data on the Web [<http://musow.kmi.open.ac.uk/>] (Daquino et al. 2018), this task will develop the *Polifonia* Web portal. It will include: 1) a reference registry of MH resources, including but not limited to all collections used and produced in the project, 2) methods for continuous indexing of MH resources, 3) methods for automatic and semiautomatic generation of metadata according to the *Polifonia* ontologies, 4) methods for searching, querying, browsing MH resources. The task will also provide input to WP6 about recommendations for the sustainability of the Web portal, beyond the lifespan of the project.

**Task 4: Preserving musical heritage through knowledge graphs** (Leader: KNAW - Participants: UNIBO, OU, KCL, NUIG, ICCD, ICBSA, Nationaal Instituut voor de Orgelkunst (NiVO) (external), Soprintendenza Archeologica Bologna (external), Soprintendenza Archeologica Genova (external))  
*This task includes two pilots focusing on building knowledge graphs for musical heritage that currently is hardly or not accessible, e.g. orally transmitted practices, embedded in texts, etc.*

**Task 5: Managing musical heritage collections through knowledge graphs** (Leader: KCL - Participants: ALL (except DP) + External adopters: BNF, CLARIN, DARIAH, CLARIAH, Europeana)  
*This task includes two pilots focusing on building knowledge graphs for supporting curators, owners, etc. in managing their large collections*

**Task 6: Studying musical heritage through (interlinked) knowledge graphs** (Leader: IREMUS - Participants: UNIBO, OU, KCL, NUIG, KNAW, Biblioteca Universitaria di Bologna (external))  
*This task includes pilots for demonstrating how *Polifonia* technologies can be leveraged to demonstrate, discover, document, support etc. research hypotheses/theories.*

**Task 7: Interacting with musical heritage knowledge graphs** (Leader: OU - Participants: UNIBO, ICBSA, The Stables Theatre, Milton Keynes, UK (external))

This task includes pilots demonstrating novel interaction solutions. The goal is to show how musical heritage knowledge graphs can be leveraged for supporting accessible and improved user experience.

## **Deliverables**

- D1.1: Roadmap and pilot requirements - 1st version (M6)
- D1.2: Roadmap and pilot requirements - 2nd version (M18)
- D1.3: Pilots development: collaborative methodology and tools (M6)
- D1.4: Intermediate validation reports for pilots: ORGANS and BELLS (M24)
- D1.5: Intermediate validation reports for pilots: INTERLINK and FACETS (M24)
- D1.6: Intermediate validation reports for pilots: TONALITIES, TUNES, MUSICBO and CHILD (M24)
- D1.7: Intermediate validation reports for pilots: MEETUPS and ACCESS (M24)
- D1.8: Final ten-pilots validation report and lessons learned (M40)
- D1.9: *Polifonia* Web portal - 1st version (M18)
- D1.10: *Polifonia* Web portal - 2nd version (M36)

## Appendix 4: Work packages in *Polifonia* project

Acronym	Title	Tasks
<b>WP1</b>	Web portal and pilots	<ol style="list-style-type: none"> <li>1. Socio-Technical Roadmap</li> <li>2. Technical coordination</li> <li>3. <i>Polifonia</i> Web portal: an aggregator of digital musical heritage collections</li> <li>4. Preserving musical heritage collections through knowledge graphs</li> <li>5. Managing musical heritage collections through knowledge graphs</li> <li>6. Studying musical heritage through (interlinked) knowledge graphs</li> <li>7. Interacting with musical heritage knowledge graphs</li> </ol>
<b>WP2</b>	Musical heritage knowledge graphs	<ol style="list-style-type: none"> <li>1. Ontology-based knowledge graphs for music objects</li> <li>2. Ontology-based knowledge graphs for music objects context</li> <li>3. Interlinking knowledge graphs</li> <li>4. Licences, ownerships, and conditions of use</li> </ol>
<b>WP3</b>	Mining musical patterns	<ol style="list-style-type: none"> <li>1. Pattern extraction</li> <li>2. Pattern recognition and definition in monodic and polyphonic music</li> <li>3. Network of musical patterns</li> <li>4. Music classification from musical patterns</li> </ol>
<b>WP4</b>	Extracting musical heritage knowledge from text	<ol style="list-style-type: none"> <li>1. Building and evaluating multilingual text corpora on musical heritage. Themes, reception, and...</li> <li>2. Automatic extraction of time, space, events, people and musical artifacts from text</li> <li>3. Automatic extraction of socio-cultural and historical context of musical heritage</li> <li>4. Evaluation of automatic knowledge extraction methods</li> </ol>

<b>WP5</b>	Human music interaction and engagement	<ol style="list-style-type: none"> <li>1. Evaluation of interaction components</li> <li>2. Multidimensional searching, browsing and exploration of musical assets and knowledge</li> <li>3. Haptic and gestural interaction with music</li> <li>4. Pattern exploration, composition and visualisation</li> <li>5. Publishing, accessing and reusing musical scholarly objects</li> </ol>
<b>WP6</b>	Dissemination, Communication and Exploitation	<ol style="list-style-type: none"> <li>1. Plan for exploitation and dissemination of results (PEDR)</li> <li>2. Web presence and <i>Polifonia</i> image</li> <li>3. Dissemination of project results</li> <li>4. Innovation drive: stakeholder network</li> <li>5. Enhanced accessibility and inclusion strategies</li> <li>6. Musical heritage promotion: the <i>Polifonia</i> open source artistic digital installation</li> </ol>
<b>WP7</b>	Project coordination and Management	<ol style="list-style-type: none"> <li>1. Overall coordination of project activities, monitoring, quality control, reporting</li> <li>2. Overall legal and contractual management</li> <li>3. Consortium agreement, management of the knowledge generated by the project and IPRs</li> <li>4. Ethical compliance</li> <li>5. Financial and administrative management and reporting</li> <li>6. Project governance and partnership communication</li> <li>7. Data Management Plan</li> </ol>
<b>WP8</b>	Ethics requirements	<ol style="list-style-type: none"> <li>1. OEI – Requirement No. 1</li> <li>2. H - Requirement No. 2</li> <li>3. H - Requirement No. 6</li> <li>4. H - Requirement No. 7</li> </ol>

		5. H - Requirement No. 8 6. POPD - Requirement No. 9 7. POPD - Requirement No. 10
--	--	---



## Appendix 5: Stories – updated template

### 1. Persona

- It is a research-based description of a typical user.
- It contains attributes such as name, age, occupation (if the Persona has more than one role, indicate which one is their primary role and which one(s) the secondary role(s)), and relevant characteristics of the person such as their knowledge and skills and their interests.

### 2. Goal

- It is a short textual description of the goal(s) that the Persona needs to be addressed in the Story.
- maximum number of characters: 1200.
- The goal(s) is(are) also represented by a short (maximum 5) list of keywords.

### 3. Scenario

- It is a Story describing how the Persona's task/need/problem is solved before, during and after interaction with the resource/software/service being developed.
- maximum number of characters: 1200.

### 4. Example Data

- This section provides examples of relationships pertaining to the Story as well as relevant links to online data

### 5. Data Resources

- List of resources (with references/links) where it is expected or known that the Persona can find what she's looking for.

## Appendix 6: Updated Survey template

### 1. DOMAIN SPECIFIC PART.

1. What are the pilot's goals?
2. What are the pilot's results and achievements so far?
3. What is the progress towards the pilot's objectives (please indicate %)
4. Are the sources digitized?
5. If so, are the sources born digital?
6. If your sources are online, can you provide links to the sources in a (partial) list?
7. Where is the data stored?
8. How scattered is musical knowledge data over different sources?
9. What is the nature of the sources?
10. Under what licences are the sources provided?
11. Who has the rights to the used datasets?
12. Research outcomes
13. What are your scientific assumptions and/or initial hypotheses?
14. What kind of scientific knowledge do you intend to produce (for example factual knowledge, conceptual knowledge, procedural knowledge, or metacognitive knowledge)?
15. Describe the knowledge units produced (melodic patterns, intertextual units, networks of composers and compositions, ...)
16. Is the new knowledge anchored to a specific fragment in a source?
17. Is the new knowledge purely descriptive? (e.g. a diplomatic transcription)
18. Can the new knowledge be the subject of dissensus?
19. Does the new knowledge involve an interpretation? And should the interpretative context make explicit through linked data?
20. Could the knowledge be inferred by algorithms, or can it only proceed from human interpretation?
21. Should it be possible to represent hypotheses and/or statements with varying degrees of certainty?
22. Could the knowledge proceed from collaborative scientific practices?
23. Does the new knowledge rely on a controlled vocabulary?
24. Does the new knowledge involve the creation of new controlled vocabularies to reflect specific concepts?
25. Could the links to the socio-cultural context be made explicit?
26. What are the pilot's main innovations? How do the methods and results contribute to the state of the art?
27. To which discipline/subdiscipline/specialty do you aim to contribute knowledge?

28. Identification/creation of analytical categories
29. Identification/creation of analytical and conceptual tools
30. Please provide links to the Competency Questions that your pilot aims to answer?
31. What are the pilot's representative Personas?
32. Please provide a specific contact person for questions related to the pilot's scientific dimension (pilot coordinator)?
33. What else has to be reported that is not covered by this part? Do you have any remarks?

**2. TECHNICAL PART.** The technical part specifies the actual implementation of the conceptual model:

1. Metadata: What formats are used?
2. Metadata: What standards are used?
3. Datasets: What formats are used?
4. Datasets: What standards are used?
5. Please list the controlled vocabularies, taxonomies, thesaurus, ontologies, etc. that you will use in your pilot
6. Please indicate if you plan to expand existing knowledge organisation systems or if you plan to create a specific one
7. Do you plan to support interoperability towards third parties?
8. Do you have APIs on your resources?
9. What algorithms will be developed for knowledge extraction or alignment?
10. Are there planned overlaps (already defined or envisioned) between the pilot's resources and other resources used within *Polifonia*?
11. What do you expect in terms of technologies from the technical provider WPs?
12. What are your needs from WP1?
13. What is the expected application to be released for this pilot?
14. Is it to be released through the Polifonia Web portal or as an independent application?
15. How should the pilot's resources be presented, linked, or stored on the website dedicated to the pilot?
16. What type of knowledge you expect can contribute to the *Polifonia* Web portal?
17. What type of interaction do you expect should be supported by the Web portal?
18. In which ways the Web portal shall link to the (separate) pilot demonstrator (if any)?
19. What kind of 'augmentation/enrichment' do you expect to get out of being displayed via the portal?

20. Can you provide a specific contact person for questions related to the pilot's technical aspects?

21. What else has to be reported that is not covered by this part? Do you have any remarks?

**3. SOCIO-PEDAGOGICAL PART.** The socio-pedagogical part aims to identify the target groups and their conditions

1. What is the target group of your pilot?

2. If academics, specify. Other?

3. Beyond the lead institution, who are the partners of the pilot and what is their involvement?

4. How is user/expert feedback to be taken into account in the pilot?

5. What kind of public demonstrators of the pilot could be made available in the coming months?

6. How do you intend to recruit users?

7. In order for the pilot to be successful, what criteria would the new tools and methods need to satisfy (e.g. certain tasks carried out more quickly or accurately, the amount or type of data that can be used in the task, number of people who are able to use the tools, the attractiveness of the interface, new tools being easy to learn, etc.)?

8. What will be the validation setup?

9. What are the pilot's validation methods?

10. What KPIs or metrics will be used for validating the pilot?

11. What values for KPIs/metrics shall be achieved for claiming success?

12. Can you name at least three stakeholders from different institutional background and countries who could contribute to the definition/validation of the pilot?

13. What facilities, if any, does the pilot make available for people with disabilities?

14. Can you provide a specific contact person for questions related to the pilot's socio-pedagogical implications?

15. What else has to be reported that is not covered by this part? Do you have any remarks?

4. Do you have any general comments and remarks?