Polifonia: a digital harmoniser for musical heritage knowledge, H2020

D1.3: Pilots development – collaborative methodology and tools (V1.0)

Deliverable information	
WP	WP1 Pilots and Web portal
Deliverable dissemination level	PU Public
Deliverable type	R Document, report
Lead beneficiary	CNRS
Contributors	OU, UNIBO, KCL, NUI GALWAY, NISV, CNAM, KNAW
Document status	Final
Document version	V1.0
Date	30/06/2021
Authors	Enrico Daga, The Open University (OU) Albert Meroño (KCL) Marilena Daquino, Fiorela Ciroku (UNIBO) Elena Musumeci (MiC) Marco Guerrieri (CNRS) Andrea Scharnhorst, Femmy Admiraal (KNAW) Raphaël Fournier-S'niehotta (CNAM)

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

PAGE INTENTIONALLY BLANK

Project Information

Project Start Date: 1st January 2021
Project Duration: 40 months
Project Website: https://polifonia-project.eu

Project Contacts

Project Coordinator

Valentina Presutti

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

E-mail: valentina.presutti@unibo.it

Project Manager

Marta Clementi

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
Research division

E-mail: marta.clementi3@unibo.it

POLIFONIA Consortium

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

Project Summary

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

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

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

Executive Summary

The deliverable reports on the collaborative methodology and tools for the technical development of the Pilots of the Polifonia Project. Technical development is coordinated by a Technical Board that designed a methodology inspired by agile software development methodologies, adapted to the needs of a research project consortium. Developers and domain experts are engaged in collaborative workshops in a co-creation process that leads to the identification of task-oriented working groups. These are developed autonomously and associated to Work Package activities. Technical outputs of the activities are collected and harmonised into a Polifonia Ecosystem - a collection of resources for musical cultural heritage preservation and reuse (software, end-user tools, data, requirement specifications, documentation, etc...). The collaborative tools for development are centred on a share space on GitHub, a Discord server for instant messaging, and a mailing list. The deliverable report on the initial work conducted on the pilots, particularly focusing on highlighting collaboration among consortium partners and shared of expertise between domain experts (musicologists, music historians) and technology experts. Finally, the deliverable illustrates preliminary plans for a Polifonia Web Portal, an *aggregator* of Musical Heritage Knowledge.

Document History

Version	Release date	Summary of changes	Author(s) - Institution
V0.1	30/04/2021	First draft released	Enrico Daga (OU)
V0.2	19/05/2021	Draft improved with contributions from partners	Enrico Daga (OU), Albert Meroño (KCL), Marilena Daquino (UNIBO), Marco Guerrieri (CNRS), Elena Musumeci (MiC)
V0.3	20/05/2021	Further improvements on the draft	Enrico Daga (OU), Fiorela Ciroku (UNIBO), Peter van Kranenburg (KNAW)
V0.4	20/05/2021	Added conclusions. Sections revised. First complete draft.	Enrico Daga (OU)
V0.5	20/05/2021	Prepared for internal review	Enrico Daga (OU)
V0.6	24/06/2021	Feedback from internal review addressed	Enrico Daga (OU), Albert Meroño (KCL), Marilena Daquino (UNIBO)
V1.0	30/06/2021	Final version	Enrico Daga (OU)

Table of contents

1 Introduction

This deliverable presents the methodology and tools for the collaborative development of the pilots of the Polifonia project. Its role is to describe plans and strategies of the the consortium for supporting the collaboration of researchers, domain experts, and developers, in designing and building the technical output of the project. Centre of the strategy is the Technical Board (Section 2). This document describes the scope and role of the Technical Board, established for coordinating the technical development and ensure the quality and consistency of the technical outputs. Crucially, we report here on the work done so far since the establishment of the TB. The first objective of the Technical Board was to define a collaborative methodology (Section 3), following the spirit of agile software engineering methodologies. Such methodology combines a two-fold approach. On the one hand, developers and domain experts are joined in sessions dedicated to specific research themes, and develop ideas on how to approach specific problems, often starting from the Pilots descriptions. On the other hand, Work Packages and Task leaders participate in those sessions with the aim of developing connections to the objectives of the project, at a broad level.

A variety of different types of assets are expected to be produced by the collaborative methodology. Those include not only software libraries, data, services, and user interfaces but also information to make sense of them: stories, representing requirements in the forms of scenarios, tutorials, and how-to. The components of the Polifonia Ecosystem (Section 4) are designed as interlinked assets for supporting the development of innovative tools for musical cultural heritage study, preservation, and exploitation. These components constitute the technical backbone of the pilots, which combine them in useful applications targeting scholars and citizens. Developers will pick and mix components of the ecosystem and compose them into complex pipelines, to build Pilot applications. However, the Polifonia Ecosystem does not start as a blank sheet. Consortium members bring expertise and technical solutions that can already be employed in such modular approach. The project has already implemented key actions to bootstrap the development activities (Section 5), including dedicated spaces on a collaborative development platform (GitHub) and a live text chat system (Discord). The methodology and tools are already employed by the working groups focusing on the development of the Polifonia Pilots, that we also summarise in this deliverable (Section 6). Finally, we report on ongoing work towards a Polifonia Web Portal, an aggregator of Musical Heritage Knowledge (Section 8).

2 Technical Board

To support the collaborative development of the Pilots (Task 1.3), the project consortium established a Technical Board (TB), composed of representative members of the partners expecting to contribute on the development of the technical outputs of the project. The role of the TB is to support the coordination of the technical activities and facilitate the interaction and collaboration. In particular, the TB monitors technical developments fostering the sharing of expertise with the objective of maximising reuse of knowledge, skills, and resources. Members of the TB are responsible of ensuring the quality of the technical outputs, both from the point of view of good practices of software engineering and from the perspective of providing guidance and support to developers in curating documentation, tutorials, and user guides. In addition, the TB supervises the curation of resources for developers, so as to maximise the reuse of the outputs by third-party organisations of the cultural heritage sector and industry. The TB establishes the methodology and tools for collaboration. The present document provides details of the methodology and tools setup so far, which will be evaluated regularly and possibly changed to adapt to the concrete needs of the pilots. Of particular importance is the dissemination of project outputs, which the TB aims at maximising by recommending the delivery of Open Source software published with a commercial-friendly Apache Licence 2.0, and asking consortium members to provide substantial arguments in case they require alternative policies.

The project coordinator appointed Enrico Daga (OU) as Technical Director (TD), whose responsibility is the coordination of the Technical Board. The currently appointed members of the TB are listed in Table 2.1. Partners are free to change the appointed person by communicating it to the Technical Director. However, the TB follows an inclusive and open approach to discussion, inviting all developers, technologists, researchers, and interested people within the consortium to join TB meetings and participate. By default, TB meetings are open to all project members and we expect to organise closed meetings for exceptional reasons only.

Table 2.1: Technical Board appointed members

Member	**Partner**	**Role / WPs or Pilots**
Enrico Daga	OU	TB director, CHILD, MEETUPS, ACCESS, WP5
Johan Oomen	NISV	WP6, INTERLINK
Raphaël Fournier-S'niehotta	CNAM	FACETS
Mathieu d'Aquin	NUIG	WP3
Rocco Tripodi	UNIBO	WP4, MUSICBO, INTERLINK, MEETUPS, BELLS
Albert Meroño	KCL	WP2, INTERLINK, FACETS
Peter van Kranenburg	KNAW	ORGANS, TUNES
Fiorela Ciroku	UNIBO	MUSICBO, INTERLINK, MEETUPS, BELLS
Marilena Daquino	UNIBO	Web portal Task Leader (WP1)
Thomas Bottini	IReMus CNRS	TONALITIES
Andrea Scharnhorst	KNAW-DANS	Leader of Data Management Task (7.1)

3 Methodology

The methodology used for the development of the pilots and the Polifonia ecosystem is inspired by agile software development methodologies [1, 2]. Similar to those, our methodology focuses on *"discovering requirements and developing solutions through the collaborative effort of self-organising and cross-functional teams and their customer(s)/end user(s)"*. This distributed, self-organised approach has several advantages, especially when confronted with a more traditional, waterfall-based model [3]:

- Does not enforce specific non-functional requirements to any of the development teams, such as programming languages, libraries, or frameworks; therefore minimising the chances of writing large, monolithic systems that are hard to document and maintain (especially after the end of the project).
- Identifies and reduces dependencies between different Pilots or WPs. By refusing to establish dependencies between the tasks of the WPs, it allows activities to progress in parallel but ensuring coherence within the framework of the project, under the WPs supervision
- It is feature-driven, and puts the requirements from the Pilots at the forefront of the development process, involving domain experts as co-creators from the start
- Allows for incremental and frequent software releases under the open-source mantra "release early, release often", to maximise the opportunity of getting feedback from the target community at early stage, and validate the output against core requirements at the early stage
- Increases decoupling and independence of components, since each component has autonomous value and can be used by and combined with other components (e.g. the Pilots and the Web Portal).
- Increases opportunity for reuse of technical and scientific artefacts
- Maximise the sustainability of the outputs beyond the scope of the project, less effective components can be abandoned, without compromising the more successful ones
- Allows for a decentralised quality assurance process, where project-global metrics can be implemented in task-specific components

Collaborative research projects such as Polifonia are confronted since the beginning with the question of how to organise the work in order to fulfil the objectives of the project and provide a unified view of the project outputs. This is typically achieved with an *integration* Work Package or Task, which has the goal of developing a unique software output, often referred to as "framework" [4]. Such artefact would incorporate WP outputs in a unified object, eventually offered to the target community as an Open Source software. However, such approach raised some concerns during the first Technical Board meeting. Particularly, the discussion converged into

questioning the utility of software frameworks in satisfying the needs of the target communities. Those concerns resonated a recent trend in software development communities[1], where it is now commonly acknowledged that knowledge sharing [5] and interaction between individuals during the process of learning [6] are the most crucial factors in software development –aspects that are typically not prioritised in "unified" frameworks . Software frameworks are monolithic in nature [7], in the sense that they prioritise the tenets of object-oriented programming above interactions between stakeholders and learning. This may impact the re-usability of its components in third-party applications, and the possibility to support applications in corner cases which were not originally considered by the framework developers. Such integrated tools provide a multiplicity of affordances in one user space reducing the ability of being effective for specific tasks[2]. In addition, providing a framework increases the learning curve to new users, which need to enter the creators' mindset to approach the solution. Finally, complex projects such as Polifonia target a variety of sub-domains, users, and scenarios. As a result, effective solutions will necessarily be heterogeneous. These concerns sketched the ground of the first Technical Board meeting (the agenda of the first meeting can be found in Appendix A), required us to approach the problem of collaborative development (and output design) in a different way.

An overview of the methodology is presented in Figure 3.1. The main ingredients of our methodology are: (a) collaborative workshops; (b) agile working groups; (c) work packages; and (d) pilots. In Polifonia, the activity is bootstrapped in *collaborative workshops*, named *Maninpasta*[3] where technology and domain experts meet to discuss themes of interest, inspired by the Pilots. The workshop starts with an open session where participants can propose relevant themes. Then, participants distribute to the various groups by following their personal interest. Often (but not necessarily), this is in line with the person role in the project. At the end of the workshop, groups report on the work done. Eventually, the workshop spins the creation of cross-functional teams that continue as independent *agile working groups*. However, this is not mandatory, and some groups discussion may not survive beyond the workshop. Often, these groups split and individual participants join other teams.

Working groups are targeted to develop specific aspects such as develop requirements of a given Pilot, or design the technical solution to solve a certain problem. working groups produce concrete outputs as components of the Polifonia Ecosystem. For example, the first Maninpasta workshop lead to the activities MusicAnnotation#1[4], MockupDesign#1[5], OntologyDesign#1[6], and BuildingKG#1[7], among others. Details of the co-creation process happening in the working groups can be found in Deliverable D1.1 - Roadmap and pilot requirements.

[1] An interesting reading on some problems with frameworks can be found at https://medium.com/@itmarketplace.net/the-problem-with-frameworks-1fafa148dbad, accessed 18/05/2021

[2] http://cloudscaling.com/blog/cloud-computing/killing-the-storage-unicorn-purpose-built-s accessed 28/05/2021.

[3] From the Italian, translates literally as "with the hands in the dough". Metaphorically, the meaning resembles the English "getting one's hands dirty".

[4] https://github.com/polifonia-project/stories/issues/12

[5] https://github.com/polifonia-project/stories/issues/10

[6] https://github.com/polifonia-project/stories/issues/7

[7] https://github.com/polifonia-project/stories/issues/6

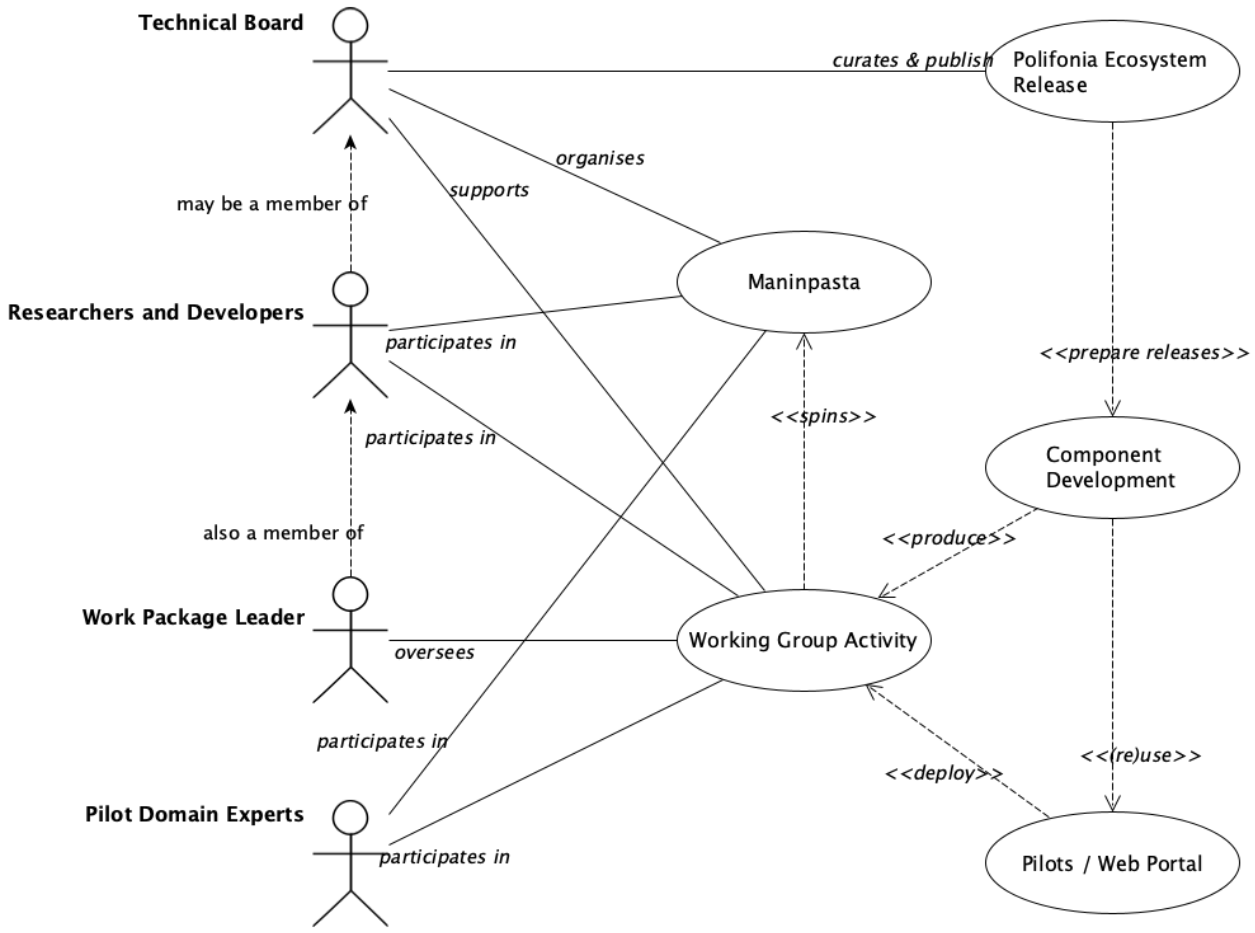

Figure 3.1: Overview of the collaborative methodology as UML use case diagram.

These cross-functional teams are organised by the various technology-oriented *Work Packages* (WP2 –knowledge graphs–, WP3 –music pattern discovery–, WP4 –text pattern discovery– and WP5 –user interfaces–). Work package leaders have the duty of matching the content of the working groups with WP tasks and organise the scientific output coherently with the WP goals.

The customers/end users are represented by the various *Pilots* (WP1), which constitute the *demand*, that the technical output under development in the working groups (the *offer*) aims at addressing. Therefore, WPs and Pilots are orthogonal and complementary views on how the consortium org. Finally, it is the role of the Technical Board to ensure the quality of the assets produced and organise them into an Ecosystem of reusable components.

Our methodology inherits several benefits from agile methodologies, while making it converge with the project framework in a two-fold way:

- **Bottom-up approach**. Component development, and especially the gathering, documentation and maintenance of requirements for such components, are managed by the working groups. This is a hackathon-like, grassroots approach that periodically ensures interaction

between requirement providers in the Pilots, and software development teams in the technological WPs.

- **Top-down approach**. WP leaders and TB members ensure that the bottom-up approach is aligned and converges, to the extent possible, to the goals established by Polifonia GA. This includes the establishment of WP checkpoints (e.g. previous to a milestone or a deliverable); and explicit breakdowns and plannings of how features, commits, releases, etc. align with the planning of WPs and their tasks. The TB synchronises and complements this top-down supervision in conjunction with the WP and Task leaders.

So, in essence, the project aims at using a bottom-up approach for gathering requirements and assembling the development teams and activity streams in working teams of domain experts (musicologists, music historians, and other target users) and technology providers (computational musicologists, MIR experts, HCI experts, web developers, and digital humanists); while we use a top-down approach to set the boundaries framing the outputs and priorities required by work-packages and concrete project goals.

In addition, our process adds the following procedures:

1. **Component development life-cycle**. Components stem from working groups and are developed in the collaborative platform (GitHub). Each component is developed in a code repository. Issue tracking and versioning ensure the development process is monitored and shared with the community that is invited to raise issues and discuss solutions. When the work towards a features is complete, the repository owner plans for a release.
2. **Sign-off release**. The release of new software components into the ecosystem must follow a code review, involving one or two code reviewers from a different team from the one that carried out the development. Specifically, the repository owner prepares a release candidate, then reviews are initiated and monitored via the Issue tracker on GitHub. When complete, the current state of the repository is marked for release, and the actual release is produced.
3. **Licencing**. Releases need to have associated explicit licencing information, which is particular important to maximise the dissemination of project outputs. The TB recommends the delivery of Open Source software published with a commercial-friendly Apache Licence 2.0. However, consortium members can choose to alternative licences, after consultation with the TB and according to the guidelines of the GA.
4. **Polifonia Ecosystem**. Components' releases are collected, annotated, and organised in the Polifonia Ecosystem, which constitutes the unified entry-point to the technical outputs of the project.

4 The Polifonia Ecosystem

The Polifonia Ecosystem is conceived as a collection of components which are both independent – they have some value on their own – and interlinked – they can be used together in order to satisfy specific end-user needs. Independence is a well known principle of software engineering, which is conceived alongside the one of inter-operability - the ability of a software component to operate with others [8]. However, the possible connections between ecosystem components don't necessarily derive from *software-to-software* relations but involve, for example, users being able to perform a complex task by using multiple tools, whose user interfaces are *linked*, or enable users to transfer data from one environment to another thanks to the mutual support of shared formats. Figure 4.1 provides an overview of the Polifonia Ecosystem, including possible relations between component types. Components types are the following (reading Figure 4.1 from right to left):

- **Registries** – indexes of resources of interest to Musical Cultural Heritage. A preliminary example is the MusoW catalogue of Musical Resources on the Web[1]. Other registries can be developed to fit specific needs (for example, the catalogue of resources useful to the CHILD pilot).
- **Ontologies** – produced in the context of the Polifonia project to support pilots and use cases, ontologies specify domain knowledge and are used as means for developing a shared understanding of the domain and as software artefacts applied in published datasets
- **Datasets** – structured data offered following best practices in (Linked) Open Data publishing. If data was not produced as Linked Data, when possible, the resource is also published in its original format. Multiple ontologies can be applied and alternative Linked Data versions of the same source data are possible, to fit the needs of different use cases.
- **Repositories and corpora** – collections of digital assets relevant to Polifonia use cases.
- **Knowledge Graph** – a distributed but unifying view of musical cultural heritage knowledge, is a *virtual* composition of Linked Data resources to be reused for large scale integration, for example, to support unified indexes for exploration and discovery.
- **Services** – Web APIs that expose reasoning and data processing capabilities. Services are run by Polifonia consortium members and instantiate specific components to the Open Web. Among those there are Linked Data services such as SPARQL endpoints – live data services publishing the above components for querying with SPARQL.
- **Software libraries** – reusable code produced by the project to support pilot activities. Software libraries are used by programmers in their own applications
- **CLI tools** – ready-made tools to be used by developers in scripting data manipulation

[1] musow.kmi.open.ac.uk

pipelines

- **User interfaces** – targeting domain experts, citizens, developed to support specific activities in the context of the Polifonia Pilots, user interfaces can be reused across similar applications targeting different datasets and scenarios
- **Experiments** – code and dataset of scientific experiments used during the research activity of the project. Experiments are meant to be documented, reproducible, and linked to research outputs. However, experiments are not expected to produce code and data which is directly reusable. The developers may produce derived assets as independent components of the ecosystem
- **Applications** – applications targeting specific use cases, possibly as direct outputs of the Pilots. Applications may reuse ecosystem components and function as demonstrators in tutorials and referenced by the documentation of the ecosystem.
- **Containers** – which wrap applications or services ready to be deployed in a computing infrastructure (a developer's laptop or a cloud service).
- **Stories** – requirements from the world of Musical Cultural Heritage preservation, exploitation, and scholarship; stories are the starting point of the collaborative methodology and the *sense-making* layer of the Polifonia Ecosystemm, giving context and purpose to the components. Stories and scenarios are linked to relevant components of the ecosystem.
- **Tutorials** – a showcase of the Polifonia Ecosystem through end-to-end tutorials, inspired from the Pilots and displaying the capabilities of the components in concrete applications. Tutorials are also an excellent starting point for developers.
- **Web Portal** (a KG view on Polifonia Linked Data resources) – the aggregator of the Polifonia Knowledge, exploiting the Knowledge Graph as underlying integration method.
- **Documentation** – documentation associated to each component. Developed autonomously, to fit the requirements and needs of the specific type of component.
- **Polifonia Ecosystem Website** (GitHub) – The resource for accessing the Polifonia Ecosystem, browsing the components and associated the documentation, accessing the resources for developers, and joining the project team in building the next generation of tools for Musical Cultural Heritage.

In component-based software design a component is a part of a system (framework), with the latter being the unifying element over the whole. However, an ecosystem as we envisage it in Polifonia, has different characteristics. Indeed, the ecosystem is a collection of artefacts, which are inherently heterogeneous, i.e. registries, ontologies, knowledge graphs, software libraries, etc. Therefore, these component types are meant to be *interlinked*. Connection points in Figure 4.1 represent the high-level strategy for components reuse and interlinking. For example, ontologies are linked to other ontologies through alignment techniques [9], and provide the knowledge representation specification for datasets (the vocabulary used in the data schema) and the knowledge graphs. Knowledge graphs, in turn, combine multiple datasets in unifying, task-oriented views. Services publish various data components and reuse software libraries which, in turn, perform computations on the datasets. Software libraries, developed and tested in Experiments, are

reused by User Interfaces, CLI tools, and Applications. In addition, Containers allow developers to deploy end-user tools easily. Scholars, citizens, and developers interact with the Ecosystem from different components. Stories illustrate the rationale behind the design of end-user tools and knowledge components, according to specific tasks and user needs. Tutorials teach users about how the Polifonia outputs can be used. The Web Portal provides access to Polifonia knowledge outputs from a unyfing view, and links to Pilot applications for in-depth analyses. Finally, a Polifonia Ecosystem Website organises the components for the benefit of developers. All Polifonia outputs are well-documented for developers to understand their functional aspects and derive opportunities for reuse. Clearly, the actors involved in the methodology presented in Section 3 are all important for the success of our strategy. However, one important role of the TB is to identify issues, for example, preventing the reuse of components, and provide guidance on how to overcome them. When appropriate, inter-operability protocols will be considered based on industry standards for the development of distributed systems on the Web. These include: W3C specifications such as RDF, SPARQL protocol, Linked Data Platform and the Solid project but also OpenAPI, JSON, JSON-LD and related technologies. For more complete references to Web Technologies and their application for metadata aggregation in cultural heritage, we refer the reader to [10, 11]. Finally, the Data Management Plan (D7.1) details management strategies and, where appropriate, long-term preservation strategies for these components[2].

However, the Polifonia Ecosystem does not start as a blank sheet. Consortium members bring expertise and technical solutions that can already be employed in such modular approach. Table 4.1 list technologies that are part of the background of consortium members and that may contribute towards supporting Polifonia applications.

[2]Some of them (e.g., registries, datasets, repositories) come with the Pilot's and are provided by institutions and groups which are member of the consortium. Others origin partly from other content or Knowlege Organisation System providers (e.g. ontologies, repositories, CLI) and will be re-used. A third part is a genuine 'product' of Polifonia itself (e.g. stories, experiments, user interfaces documentation). It is this last category for which the Data Management Plan details specific actions.

Table 4.1: Technologies of consortium members that may contribute to the Polifonia Ecosystem

Name	Links	Type	Provider	Relations to WPs / Pilots / Notes
LED	http://led.kmi.open.ac.uk Data available at http://data.open.ac.uk/sparql	Dataset, Service	Enrico (OU)	Mainly related to CHILD, MEETUPS, and WP4
MIDI LD	https://midi-ld.github.io	Dataset	Albert Meroño (KCL)	Basis for WP2, useful for WP3. Large RDF KG and ontology of linked MIDI file contents from 500K MIDIs from the Web
SPARQL Anything	https://github.com/SPARQL-Anything/sparql.anything	Software library, CLI tool	Enrico Daga (OU)	Support tool to re-engineer non-RDF resources into Linked Data. Supports CSV, JSON, XML, HTML, ...
RAMOSE	https://github.com/opencitations/ramose	Software	Marilena (UNIBO)	Python API manager on top of SPARQL endpoints
Lucinda	https://github.com/opencitations/lucinda	Software	Marilena (UNIBO)	RDF browser based on JSON templates
OSCAR	https://github.com/opencitations/oscar	Software	Marilena (UNIBO)	RDF Search engine based on JSON templates
MusoW	https://musow.kmi.open.ac.uk/	Linked Data	Enrico Daga (OU)	LOD registry of MH on the web to be expanded / enriched
ArCo	https://w3id.org/arco	Linked Data / software	Val (UNIBO)	Ontology network and LOD / XML2RDF for normative italian Cultural Heritage XSD
Lizard	https://github.com/anuzzolese/lizard	Software	Valentina Presutti (UNIBO)	Automatic generation of ontology-based APIs for querying knowledge graphs
Neuma	http://neuma.huma-num.fr	Platform: library/dataset	Raphaël FS (CNAM)	WP1: Pilot FACETS MEI corpus, REST API, vizualisation of scores, analysis/annotations.
Framester	https://github.com/framester/Framester	Linked Data / Software	Fiorela Ciroku (UNIBO)	Frame-based ontological resource
Edwin	https://github.com/luigi-asprino/edwin	Framework	Fiorela Ciroku (UNIBO)	Builds and analyses Equivalence Set Graphs
Cultural-ON	https://dati.beniculturali.it/cultural-ON/ENG.html	Linked Data	Fiorela Ciroku (UNIBO)	Data on cultural institutes or sites
Multilingual Corpus LBC	http://corpora.lessicobeniculturali.net/en/	Dataset	Fiorela Ciroku (UNIBO)	MusicBo
Unicittà Corpus	Available in July 2021	Dataset	Fiorela Ciroku (UNIBO)	MusicBo
CultuurLINK	https://cultuurlink.beeldengeluid.nl/	Software	NISV	Open Source Tool for aligning vocabularies
Spinque Desk	https://spinque.com/	Software (commercial)	NISV/ Spinque	Integrate your data into a knowledge graph, design search solutions tailored to your needs and deploy them as APIs.
FindLEr	https://github.com/enridaga/led-discovery	Software	Enrico Daga (OU)	Supports the retrieval of listening experiences in digitized books.
grlc	http://grlc.io https://github.com/CLARIAH/grlc	Software (OS)	Albert Meroño (KCL)	Supports automatic creation of KG APIs from shared SPARQL queries
CLOVER	http://arco.istc.cnr.it:8081/	Software	Valentina Presutti (UNIBO)	A prototype instance of OntoPortal for CH ontologies
SPICE Linked Data Hub	http://spice.kmi.open.ac.uk http://github.com/mkdf/	Software (OS)	Jason Carvalho (OU)	Example instance of a Linked Data Hub developed using the MK Data Factory suite. Supports data management (file repository, JSON streams), transformations to RDF (via SPARQL Anything) and publishing of SPARQL endpoints.

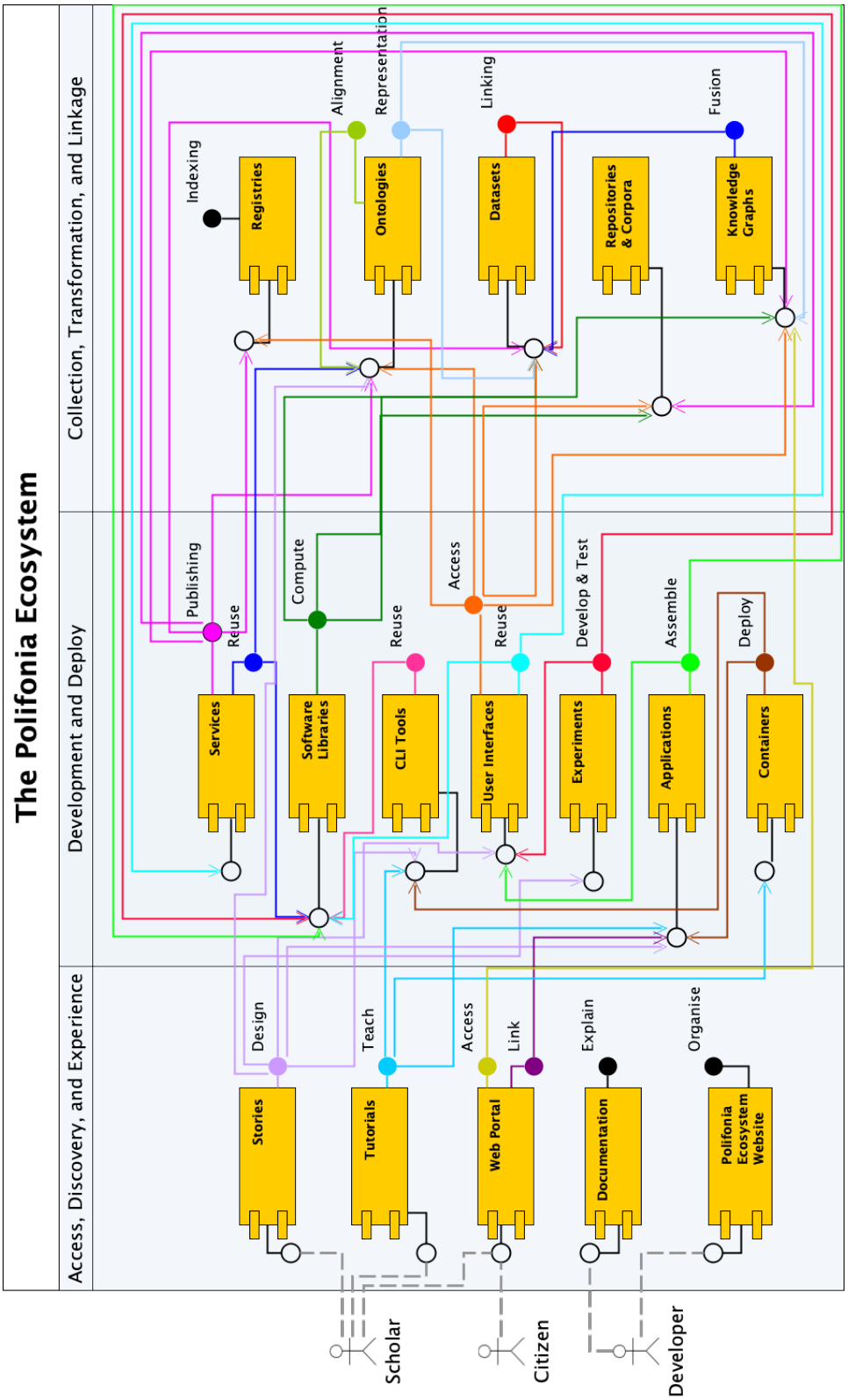

Figure 4.1: The Polifonia Ecosystem: overview of the components and main relations. Connection points in black represent connections to all the components.

5 Collaboration Infrastructure

The Technical Board setup the following infrastructure to support the collaborative development of the Pilots.

GitHub is the leading code hosting platform supporting development teams in tasks such as version control and collaboration. The TB established the **polifonia-project** GitHub organisation[1] as the collaboration space for the technical activities of the consortium. The GitHub organisation currently includes 30 registered GitHub user accounts belonging to contributing project members. The space is used as the primary infrastructure for supporting collaborative development of the pilots, including collection of requirements, software development, testing, and documentation. A key element of the approach include an *open-by-default* policy, where contributions to the development is encouraged also from outside the Polifonia consortium. So far, the collaborative space was used to collect stories and scenarios, including managing the discussion and actions of activity streams emerged in the Maninpasta workshop. Repositories include: **stories** – which collects requirements and issues emerging during the first two Maninpasta workshops, **registry** – which is dedicated to the development of a Polifonia Registry of musical resources, from the original work of [12], and **ecosystem** – aimed at hosting the Polifonia Ecosystem Website.

Discord is a platform supporting invite-only social media. The TB created a Polifonia Server that is currently used by project members for fast interaction. Communication is organised in channels:

- general – for general discussions
- dev – for technical discussions of developers and the technical board
- maninpasta – for live interaction during the workshops
- wp1, wp2, wp3, wp4, wp5 – for discussion relevant to specific WPs and for live interaction during periodic update meetings

Consortium members, task leaders, and working groups are free to create channels dedicated to specific activity streams. Currently, users created the following: *sethusmockup*, *textcorpus*, and *ontologyengineering*.

Finally the infrastructure offered by the coordinator (UNIBO) – via Microsoft Sharepoint – includes a Technical Board **mailing list** that is open to all project members interested in following or contributing to technical activities. Developers are particularly invited to subscribe, join technical board meetings, and use the list for technical discussions.

[1] http://github.com/polifonia-project

6 Pilots development, so far

This section provides an introduction to the Polifonia pilots, discussed in the context to the collaborative methodology and tools.

In the DMP for each Pilot 'data stories' are written which describes them according to the ORDP Data summary section, addressing questions such as: what is the purpose of the data collection and its relation to the objectives of the project; what types and formats of data will be used; how much data are re-used; what is the origin of the data; its size, and its envisioned use. The DMP also contains an Appendix listing factual information about the data in the Pilots. Here, we zoom into the Pilots from the perspective of their contribution to the envisioned final knowledge graph.

[ORGANS] - A Knowledge Graph on History of Pipe Organs Technology provider(s): WP2, WP4. Beneficiaries: KNAW, NIvO(external).

Pilot objectives: The objective of the pilot is to derive a knowledge graph from the full contents of the Organ Encyclopedia.

Work thus far: The main work on the pilot will start per September 2021. Some preliminary tasks have been done, including: a) Assembled the files containing the full text of the Organ Encyclopedia; and b) Interviewed stakeholders to collect potential use cases. Discussion are ongoing between KNAW and the TB coordinator on possible approaches to the extraction of content from the Word documents.

[BELLS] - Preservation of Historical Bell Heritage: dependencies between tangible and intangible Technology provider(s): WP2, WP3, WP4. Beneficiaries: MiC –Ministry of Culture (ICBSA-Central Institute for Sound and Audiovisual Assets; ICCD-Central Institute for Cataloguing and Documentation)

Pilot objectives:The widespread presence of bell structures constitutes a cardinal element of the landscape in Italy, contributes to the definition of a soundscape, performs a function of marker of the daily and festive / ritual time. The objective of the project is the representation of this phenomenon through the realization of a database of existing sources through music and text extraction technologies (WP4,WP4) and the construction and representation of surrounding knowledge through ontology modeling and the realization of a knowledge graph (WP2). This will help the analysis of the bell heritage to the wider context of a landscape and cultural heritage, both in its tangible and intangible aspects, and will be addressed to researchers, institutes that deal with the construction of information systems for the protection and enhancement of cultural heritage, local public bodies and other local actors, engaged in the protection, safeguard, use and enhancement of cultural heritage, also for the development and definition of landscape plans. The Pilot team

is very heterogeneous: musicologists and ethnomusicologists, archivists, ethnoanthropologists, architects.

Work thus far:The following activities have been carried out so far: elaboration of a scenario on GitHub relating to the needs of restoration and conservation in compliance with the performance and sound practices needs [https://github.com/polifonia-project/stories/Keoma:Architect]; analysis of the scenario and competency questions with ontology design experts during the first Maninpasta hackathon; identification of useful sources for the construction of a corpus of texts with linguistic experts during the second hackathon; elaboration of a second scenario on GitHub focusing on intangible heritage safeguarding practices [https://github.com/polifonia-project/stories/tree/main/Patrizia:ethnoantrhropologist]; development of a workflow for interviews and digitization of existing sound resources together with local Superintendences; identification of local partners (bell ringers associations and local experts in campanology); discussion about the development of a bottom-up process of cooperation for the identification of vocabularies about informal practices for the transmission of knowledge about sound practices and techniques.

[INTERLINK] - Interlinking of collections in digital music libraries and audiovisual archives Technology provider(s): WP2, WP3, WP4. Beneficaries: KNAW, NISV, ICCD, ICBSA, UNIBO, IREMUS, NUIG; CLARIN, DARIAH, CLARIAH, Europeana (external)

Pilot objectives: INTERLINK seeks to connect collections in digital music libraries and audiovisual archives in a meaningful way by means of a Knowledge Graph. Within the Polifonia project the INTERLINK-pilot establishes the required infrastructure for the analysis of relations between musical heritage across different collections. By drawing on knowledge graphs (WP2) and music and text extraction technologies (WP3, WP4), this pilot will explicitly reveal and make compatible the entities and concepts hidden in digital music libraries and audiovisual archives.

Work thus far: A user story, 'William'[1], was drafted specifically for the INTERLINK pilot, in which specific attention was paid to the need for a Knowledge Graph based on catalogue metadata. More in depth analysis of pitch, rhythm and other modalities of music could become a further part of the Knowledge Graph, after the more basic catalogue metadata has been linked. The musoW platform[2] functions as a registry for various music collections. Discussions are ongoing about how this platform can be improved upon, in terms of the data format used and the ease with which people can contribute new datasets to the platform. Finally, the option of using GitHub as the database for the registry is being assessed.

[FACETS] - Exploration of music scores collections through statistical features Technology provider(s): WP3, WP4, WP5. Beneficaries: CNAM, Iremus, NUIG, BNF (external).

Pilot objectives: FACETS seeks to improve exploration and discovery of large collections of scores through the creation of a faceted search engine (FSE). It will rely on features extracted

[1]See https://github.com/polifonia-project/stories
[2]https://musow.kmi.open.ac.uk/

and identified in WP3 and WP4 (melodic, harmonic or rhythmic patterns, style, structure, instrumentation, metadata). This engine will be demonstrated in Neuma, and its code released in open source. Other score-oriented musical libraries for cultural heritage will benefit from the code, such as Royaumont and Gallica-BNF. Results will be reused by WP2 and WP3.

Work thus far: A PhD student, Tiange Zhu, started her work mid-February. She has so far worked towards getting familiar with the Neuma platform[3] and extending preliminary works on a search engine dedicated to musical scores. The search engine currently features an exact melodic search, or a transposed one. A rhythmic search and a lyrics search are developped, as well as some refinements for melodic patterns ("mirror search"). The migration of the source code on Github will happen in the next few months, and a journal paper will be submitted in a similar schedule (extending the results of [13]).

[TONALITIES] - Modal and tonal classification of Western notated music from the Renaissance to the 20th century **Technology provider(s)**: WP2, WP3, and WP5. Beneficiaries: musicians, students, academics (mainly digital musicologists and music analysts), amateur internet users.

Pilot objectives: This pilot investigates modal-tonal identification, exploration and classification of monophonic and polyphonic notated music from the Renaissance to the 20th century.

Work thus far: corpus identification; design of a general scheme of software components; granular identification of potential components and deliverables; functional specifications sketched; progress in the definition of specific ontologies/controlled vocabularies; first attempts/tests of data entry interface; proposition of four stories relating to Tonalities, one of them (Sethus) chosen by Polifonia members as case-study and analysed/developed during the Maninpasta/hackathons; starting contacts and exchanges with the MEI community.

[TUNES] - Tunes analysis and classification **Technology provider(s)**: WP2, WP3. Beneficiaries: KNAW, Utrecht University.

Pilot objectives: Link the full contents of the collection of Dutch Song data of the Meertens Institute with other European collections of traditional and composed music.

Work thus far: The main work on the pilot will start per September 2021. Some preliminary tasks have been done: a) Interviewed stakeholders to collect potential use cases; b) Initial discussions on the types of patterns that will be used to interlink the various collections.

[MUSICBO] - Knowledge graph of Bologna Musical Heritage **Technology provider(s)**: WP2 and WP4

Beneficiaries: UNIBO, Biblioteca Universitaria di Bologna (external).

Pilot objectives: To identify evidence of the role of Bologna as a creative city for music over centuries.

[3]Developed with Python/Django, see http://neuma.huma-num.fr

Work thus far: In order to reach the objective of the pilot, we are creating a knowledge graph of the Bologna Music Heritage by aggregating data from different sources into a unifying view. The data is sourced from Lessico Beni Culturali, the BUB archives, as well as other open access libraries. The work consists on gathering a chronology of the main musical events in Bologna covering a period from 1400 to 1900 and a list of names of key figures in the musical scene in Bologna. At this point in time, the team dedicated to this pilot is creating a balanced and wide corpus. Also, a preliminary map of digital repositories with an indication of the texts they have, e.g. Museo internazionale e biblioteca della musica for letters of padre Martini, is being developed.

[CHILD] - Exploration of musical heritage for scholarly enquiry: a case study on Music and Childhood Technology provider(s): WP4, WP5. Beneficiaries: Musicologists and historians of music, teachers, citizens.

Pilot objectives: Exploring a historical perspective on the part music has played in children's lives through education, play and family and community practices, and how far such experiences differ across time, culture and gender. Supporting music scholars from the formulation of a hypothesis to the discovery, collection, and curation of resources relevant to the enquiry.

Work thus far: exploration of the Listening Experience Database, Archive.org, and Gutenberg Project is under development with the goal of defining a corpus/registry of sources potentially relevant to the study. Analysis of requirements developed in interviews with domain experts and as part of the Maninpasta workshop. Stories and scenarios relevant to the pilot were developed (see the scenario Ortenz - Music and childhood[4])

[MEETUPS] - Musical Meetups: the European musicianship flow Technology provider(s): WP4, WP5. Beneficiaries: musicologists and music historians (OU, UNIBO, and ICBSA).

Pilot objectives: 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. Objective of the pilot is providing a Web tool that extracts relevant information from public domain texts and a user interface for mapping encounters between people in the musical world in Europe, c.1800-c.1945.

Work thus far: activities focused on requirement analysis for the pilot. The Maninpasta workshop and independent activity streams produced three scenarios relevant to the pilot: Ortenz - Musical

[4] https://github.com/polifonia-project/stories/blob/main/Ortenz:%20Music%20Historian/Ortenz%20-%20Music%20and%20childhood.md

social network[5], Sophia – Musicians and their environment[6], and David - Music Historian[7]. Ongoing work includes surveying existing databases which may provide the starting point for building a dataset of meetups, considering sources such as Wikipedia (DBpedia, Wikidata) and Linked Data resources such as the EventKG as well as ontologies and schemas for Event representation. Other datasets and collections involved for gathering sources of events are Archive.org, Gutenberg Project, Biblioteca Italiana, and BNF France.

[ACCESS] - Making musical performances accessible to people who are Deaf or hearing impaired Technology provider(s): OU

Pilot objectives: The aim of this pilot is to support people who are Deaf or hearing impaired in participating as audience members in musical performances. The work will focus on the development of socio-technical solutions involving worn haptic technologies relaying properties of the musical performance in real time.

Work thus far: An initial story for the pilot has been developed which can be found on the GitHub repository. This involves haptic bracelets relaying different musical properties which can be worn, swapped and shared by participants during a performance. Current haptic technologies are being reviewed for their suitability. As part of this review, we are collaborating with the MIRAGE project, funded by The Research Council of Norway, which is developing computational tools for the analysis of musical properties such as rhythm. Discussions have commenced with The Stables (early adopter) which has only just reopened following lock-down.

[5] https://github.com/polifonia-project/stories/blob/main/Ortenz:%20Music%20Historian/Ortenz%20-%20Musical%20social%20network.md#ortenz---musical-social-network

[6] https://github.com/polifonia-project/stories/blob/main/Sophia:%20Musicologist/Sophia%23MusiciansAndTheirEnvironment.md

[7] https://github.com/polifonia-project/stories/blob/main/David:%20Music%20Historian/David%231_musichistorian.md

7 Web portal: an aggregator of digital musical heritage collections

The Polifonia Web portal is designed to be the main access point to data collections produced in the pilots. The objectives of the portal are threefold: (1) to provide user-friendly interfaces and let the general public access valuable information on musical heritage, (2) address exploratory and analytical tasks targeted on specific scholarly groups, and (3) provide access and foster reuse of data sources.

The design and development of the Web portal is iteratively informed by activities carried out by working groups, namely:

- Data layer requirements (WP1, WP5, TB)
- Software solutions (WP7, TB)
- Infrastructure requirements (TB, project coordinator)
- Sustainability plans (TB, project coordinator (via DMP-ORDP), third-parties)

Data layer requirements. The socio-technical roadmap (see Deliverable D1.1) offers an overview of *contents* relevant to pilots. It identifies shareable metadata, multi-modality aspects, and linking between data sources to be produced by project members and sources already available on the web (see also the Data Stories and Annex of D7.1. First version of the Data Management Plan). The roadmap provides us with the big picture of content requirements, possible overlaps between data created by pilots, and whether the development and access to certain data sources must be prioritised.

Interviews with scholars participating in music research are carried out to by OU and UNIBO in order to frame the following aspects: (1) data-driven research questions, (2) *data-sense making activities*, and (3) metadata requirements. Domain experts involved in Polifonia pilots are asked to detail their research methods, whether sources are already digitised or not, and whether computer-aided tasks are in scope. Content analysis of transcribed interviews is performed to frame research methods into sense-making primitives and define requirements of the final web interfaces. The latter include tools for exploratory data visualisation (EDA), features of the search engine, and User eXperience (UX) aspects.

Software solutions. Software solutions that are part of Polifonia ecosystem (see section 4) are periodically surveyed by members of the technical board to frame the state of the art, and if needed, propose a intervention of Polifonia developers to fill the gaps. Software solutions include, among the others, (1) *ontologies and vocabularies* that Polifonia data sources rely on (data layer), (2) *software libraries* for Web development, data processing, and Linked Open Data manipulation, and (3) *standards* for music annotation.

A working group including representatives of pilots and the TB is dedicated to the development and maintenance of a *registry* of musical resources on the web. The work extends and optimizes the existing musoW registry[1]. The objectives of this activity are the following: (1) to develop sustainable solutions for updating and maintaining the online registry with metadata about data sources produced by the project, (2) to allow crowdsourcing of curated metadata of other data sources and tools relevant to Polifonia objectives and (3) to provide the final web portal with a strategy for harvesting up to date data collections. Crowdsourcing allows us to discover interesting resources, to annotate them with contents that cannot be automatically extracted (e.g. references to historical periods or characters addressed in the documents), and to support initial mechanisms of interlinking and discovery across collections. The development of crawling mechanisms to collect, update, and store crowdsourced contents will guide the development of reusable solutions for dynamically harvesting contents from data collections (indexed in the registry) that may change over time.

Lastly, this activity is aligned with the Data Management Plan, which details information about the openness and FAIRness of the components of the Polifonia Ecosystem, and it is informed by the Dissemination plan (see D7.1. which details data publication and dissemination strategies). Such strategies are iteratively refined according to data volume and long-term preservation purposes, and affect the way the web portal accesses the data layer.

Infrastructure requirements. The TB defines suitable hosting and deployment solutions to guarantee optimal usage of available resources. We expect infrastructure requirements to vary along with the evolution of data layer requirements. Our work is inspired by previous works addressing similar situations [14, 15], namely: distributed systems leveraging data sources stored in different locations, systematically updated by providers, and served according to different access and licensing options.

Fig. 7.1 illustrates the high-level infrastructure of the mid-term/final version of the web portal and the dependencies with Work Packages that are involved.

As shown in the picture, we expect data sources to be shared by partners and external data providers in different ways, including SPARQL endpoints, REST APIs, Linked Data Fragments, and RDF data dumps. Among the sources, the musoW catalogue acts as a centralised registry recording cataloguing data of sources that will populate the web portal.

The web portal back-end includes a dedicated triplestore, which stores selected information harvested from data sources in the form of indexes and linksets. Indexes include selected metadata of most recurring entities across pilots (e.g. music works, people, instruments), so as to quickly retrieve data requested by searching and browsing interfaces. The selection of indexed metadata aims to address requirements of most web portal views, and it is informed by activities grouped under *Data layer requirements*. Linkesets include links to external open datasets (e.g. Wikidata, DBpedia) that facilitate interlinking between pilots data sources - wherein the same entities may be differently identified (e.g. different IRIs to identify the same person) - and allow us to serve precise, complete, query results. Specialised services perform pre-processing (cleaning), reconciliation (deduplication), and transformation activities (according to a crawling schema).

[1] https://musow.kmi.open.ac.uk/

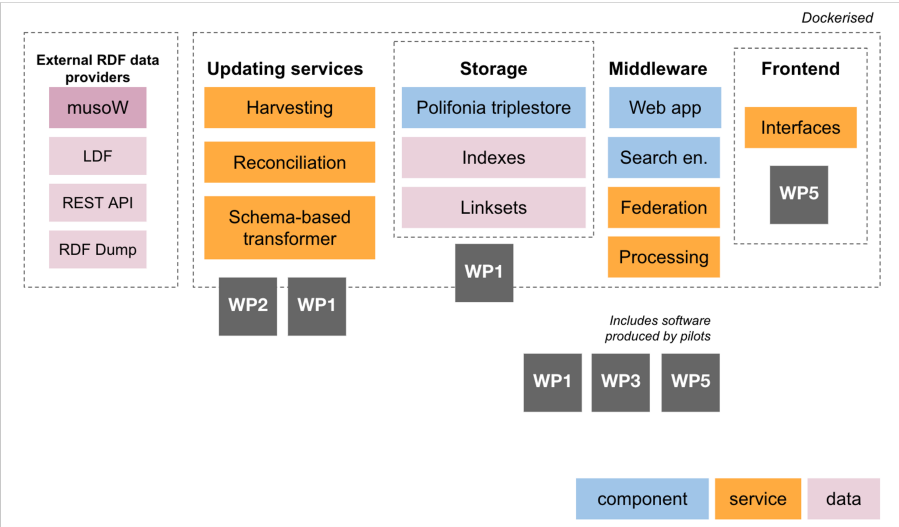

Figure 7.1: Overview of Polifonia Web Portal infrastructure

The middleware consists of a fully compliant Model-View-Controller (MVC) application, which handles the search engine, federated queries to external sources (when applicable), post-processing of data, and serves data as views.

To deploy and deliver software quickly we use Docker.[2]

Sustainability plans The project coordinator investigates sustainability plans for the Polifonia web portal, with support from the TB. In particular, options for *long-term access, hosting, and maintenance*, of the web portal and data sources may involve third-parties interested in contributing with trusty solutions. Webinars around the Data Management Planning will inform the consortium about the state of the art of Fair Data, Fair Digital Objects and preservation strategies. The next iterations of the Data Management Plan (version 2 and 3) will document the final decisions made concerning mid- and long-term sustainability.

It is worth noting that one of the advantages of relying on Linked Open Data and Semantic Web technologies includes the usage of *persistent, dereferenceable HTTP URIs* for identifying resources, web documents, and data collections. That is, a end user will always be able to access web resources by using the same persistent URIs regardless their actual location and storage solutions. So doing, we plan to iteratively increment infrastructure requirements without affecting data access and reuse strategies.

[2] https://www.docker.com

8 Conclusions

In this deliverable we reported on the collaborative methodology and tools for the development of the Polifonia Pilots. This methodology is based on a collaborative solution that integrates and orchestrates within a single agile development model different actors (i.e. the technical board, working groups, customers, and users), activities (i.e. collaborative workshops, aka *maninpasta*, WP tasks), and artefacts (i.e. components of the Polifonia ecosystem and the ecosystem itself). The organisation of the development activities follows the principle of well-established agile software development practices but also specialises them in relation to the needs of a research project consortium. Additionally, the deliverable describes the nature of the components part of the Polifonia ecosystem (i.e. registries, datasets, knowledge graphs, software libraries, etc.) and identifies a tentative list of possible technologies that can be adopted for realising the ecosystem of the Polifonia project. The methods and tools are already used by the consortium members in the several activities, as reported in the last sections of the deliverables. Decisions taken so far are based on past experiences of technical board members in similar large scale research projects. Although we are confident that an agile methodology and a loosely-coupled approach to product development (the ecosystem) has several benefits compared to a water-fall approach and a tight software integration (the framework concept), the TB is committed to reviewing the methodology periodically, and adapt the decisions to better meet the project goals as well as the experience of the developers. Considering the incremental nature of the work, immediate next step include the early stage releasing of specific components deriving from background technologies (e.g. a registry of musical resources derived from musoW[1]) and the detailing of the components development life-cycle, including a better understanding of roles and responsibilities of the actors involved in it.

[1] http://musow.kmi.open.ac.uk

Bibliography

[1] K. Collier, *Agile analytics: A value-driven approach to business intelligence and data warehousing*. Addison-Wesley, 2012.

[2] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, S. Mellor, K. Schwaber, J. Sutherland, and D. Thomas, "Manifesto for Agile Software Development," http://agilemanifesto.org, 2001.

[3] H. D. Benington, "Production of large computer programs," *Annals of the History of Computing*, vol. 5, no. 4, pp. 350–361, 1983.

[4] R. E. Johnson, "Frameworks=(components+ patterns)," *Communications of the ACM*, vol. 40, no. 10, pp. 39–42, 1997.

[5] J. Kalermo and J. Rissanen, "Agile software development in theory and practice," 2002.

[6] K. Suryaatmaja, D. Wibisono, A. Ghazali, and R. Fitriati, "Uncovering the failure of Agile framework implementation using SSM-based action research," *Palgrave Communications*, vol. 6, no. 1, pp. 1–18, 2020.

[7] D. Riehle, "Framework design: A role modeling approach," Ph.D. dissertation, ETH Zurich, 2000.

[8] I. Crnkovic and M. P. H. Larsson, *Building reliable component-based software systems*. Artech House, 2002.

[9] J. Euzenat, "An api for ontology alignment," in *International Semantic Web Conference*. Springer, 2004, pp. 698–712.

[10] N. Freire, G. Robson, A. Isaac, H. Manguinhas, and J. B. Howard, "Web technologies: A survey of their applicability to metadata aggregation in cultural heritage." in *ELPUB*, 2017, pp. 235–244.

[11] E. Daga, L. Asprino, R. Damiano, M. Daquino, B. Diaz Agudo, A. Gangemi, T. Kuflik, A. Lieto, M. Maguire, A. M. Marras, D. Martinez Pandiani, P. Mulholland, S. Peroni, S. Pescarin, and A. Wecker, "Integrating citizen experiences in cultural heritage: requirements, state of the art, and challenges," *(to appear in) Journal of Computing and Cultural Heritage*, 2021.

[12] M. Daquino, E. Daga, M. d'Aquin, A. Gangemi, S. Holland, R. Laney, A. M. Penuela, and P. Mulholland, "Characterizing the landscape of musical data on the web: State of the art and challenges," 2017.

[13] P. Rigaux and N. Travers, "Scalable searching and ranking for melodic pattern queries." in *ISMIR*, 2019, pp. 801–808.

[14] F. Kirstein, S. Dutkowski, B. Dittwald, and M. Hauswirth, "The european data portal: Scalable harvesting and management of linked open data." in *ISWC Satellites*, 2019, pp. 321–322.

[15] F. Kirstein, K. Stefanidis, B. Dittwald, S. Dutkowski, S. Urbanek, and M. Hauswirth, "Piveau: A large-scale open data management platform based on semantic web technologies," in *European Semantic Web Conference*. Springer, 2020, pp. 648–664.

Appendix A. Agenda of the first Technical Board meeting

A summary of the agenda of the first Technical Board meetings, defining the issues driving the scope and goals of TB activities in Polifonia.

1. Technical board: scope and role
    a) Coordinate technical activities / Facilitator
    b) Share technical expertise
    c) Maximize reuse
    d) Ensure quality of outputs / documentation
    e) Support dissemination of outputs (GH web site, documentation, tutorials)
    f) WP and Pilots are Orthogonal
        i. identify dependencies
        ii. Increase opportunity for reuse (scientific / technical)
2. The Polifonia Ecosystem: what it is (not) and what is the role of the TB
    a) Components types
        i. Interoperation strategies: Web technologies (OpenAPI, SPARQL, Solid, URLs, . . .)
        ii. Datasets: Ontologies, published following best practices; SPARQL endpoints; Versioned Releases (on GitHub) ;
        iii. Software libraries: Code quality; Documentation; Tutorials; CLI.
        iv. Services / Web APIs: Availability; Good practices (e.g. OpenAPI specification); Documentation; Tutorials.
    b) User Interface components
        i. Good practices:
            A. Linkable UI views
            B. Reusable snippets (Web Embed)
        ii. Reusability of UI
    c) Web Portal: what is the role of the WP in the ecosystem?
    d) Demonstrators
3. Development Methodology
    a) Agile, feature-driven
    b) Incremental releases: "release early, release often”

   c) Decoupling / independence (each component has an autonomous value but *can* also be used with others, as exemplified in Pilots and Web Portal)
   d) Quality assurance (strategy?)
   e) Semantic Versioning
   f) Registry of components (GitHub repositories?)
4. Development Team: All members of the GH organization
5. Collaboration Infrastructure
   a) GitHub
   b) Discord
   c) Mailing list of the TB also for the development team
6. Milestones Deliverable at M6 should define:
   a) Methodology -> coherent with Socio-technical roadmap
   b) Guidelines for developers (cookbook) -> coherent with DMP
   c) The Ecosystem at a glance (component types and interaction methods)
   d) Plan, timeline, etc...
7. Next actions
8. Develop a plan for the technical developments of Y1
9. Develop a timeline of component releases towards Y1
10. Next meeting
11. AOB