

## The Polifonia Research Ecosystem: an Executable Data Management Plan

Authors: Enrico Daga, Andrea Scharnhorst, Raphael Fournier-S'niehotta, Marco Gurrieri, Jacopo de Berardinis, James McDermott, Marilena Daquino, Jason Carvalho, Marco Ratta, and Christophe Guillotel-Nothmann

This paper introduces an innovative approach to documenting research components (including research data). The *Research Ecosystem* approach (Daga et al. 2023) gives guidelines to determine which components are relevant in the light of certain research questions, how to annotate them as semantic, machine-readable artifacts, and how to validate, control and preserve them. Developed in the context of Polifonia<sup>1</sup>, a semantic-web-based research project to improve access to musical cultural heritage, the framework is re-usable in all projects which rely on collaborative, open software development.

We zoom into specific challenges which emerge when dealing with digital objects from the cultural heritage domain - in our case music. In particular, we showcase how to achieve machine-readable expressions of license information, and enrichment of metadata supported by Large Language Models.

Ultimately, the Research Ecosystem fosters the management of Data in their context, namely together with Tools and Reports. However, it goes beyond pure documentation. In the case of Polifonia, archetypical users with their specific information needs (described as Personas and Stories) constitute one important component type. Components are connected via annotations. This way, the digital-humanities-born specific methodological approaches, the data used, the tools built and the documentation produced, all form a network which shows the interdependencies of research questions, methods, and data. In this component-based structure, traditional project management elements such as Work packages and Tasks also appear as part of the annotation scheme.

The beauty and efficiency of the approach lies in the consequent use of existing platforms such as GitHub<sup>2</sup>, which already contain elements to build a machine-based Ecosystem. One result is a machine-executable Research Data Management plan, which takes standardization in RDM to a next level.

The challenge in re-using the Ecosystem approach is to find answers to classic research management questions such as what are the right components, which annotations are needed to express the interdependencies, and how both components and their links need to change over time. To support re-use of concept, Polifonia defined workflows for building and evolving an Ecosystem next to workflows which form part of the Ecosystem. This way, documentation and management are intrinsically interwoven with classical research management questions: how to find the right questions, how to find the right methods, and how to organize collaboration in an interdisciplinary setting (Guillotel-Nothmann et al., 2022).

Our Research Ecosystem approach builds on other approaches to formal description of research assets (such as FAIR Digital Objects, or RO-CRATE), but its essence is to address a middle level: above the elements of research but below the large-scale research objectives. We

---

<sup>1</sup> <https://polifonia-project.eu/>

<sup>2</sup> <https://github.com/polifonia-project>

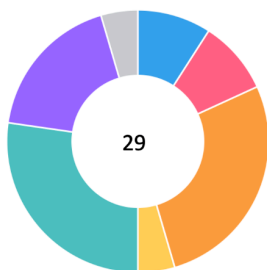
will show how the Polifonia Ecosystem evolved, and how the Ecosystem improves efficiency when tracing the quality and FAIRness of the research output and processes.

This paper addresses in particular the following questions from the call: What is the role of interdisciplinarity? What does it mean for a workflow to be ethical, reproducible and sustainable? What kind of documentation is necessary and at what level of granularity? Are there data management frameworks that make the documentation of workflows easier?

#### Data

This section collects project outputs that are released as *data*. The ecosystem considers data any digital object that specifies, describes, or represents facts about the project's domain of interest. These include various types of digital objects such as *datasets, corpora, ontologies, or repositories*.

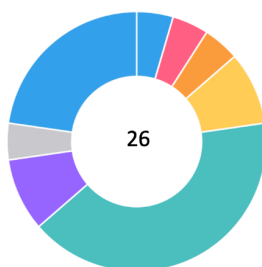
Corpus Dataset KnowledgeGraph Lexicon Ontology Repository  
Scheme



#### Tools

This section collects project outputs that are released as *software or application*.

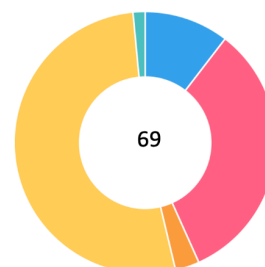
Application CLTool DockerImageContainer SPARQLEndpoint Software  
SoftwareLibrary UserInterface WebApplication



#### Reports

This section collects project outputs that are released as *report*. The ecosystem considers reports any digital object that specifies, describes, or represents facts about the project's domain of interest. Reports differ from *data* as they are mainly directed to human consumption, rather than computational treatment. Reports include various types of digital objects such as *documentation, tutorial, requirements collections, stories or persona specifications*.

Documentation Persona RequirementsCollection Story Tutorial



The figure shows visual overviews from the Ecosystem website<sup>3</sup>.

## References

Daga, E., Daquino, M., Fournier-S'niehotta, R., Guillotel-Nothmann, C., & Scharnhorst, A. (2023). Documenting the research process. Opportunities and challenges for Bibliometrics and Information Retrieval. In I. Frommholz, P. Mayr, G. Cabanac, S. Verberne, & J. Brennan (Eds.), Proceedings of the 13th International Workshop on Bibliometric-enhanced Information Retrieval co-located with 45th European Conference on Information Retrieval (ECIR 2023) (pp. 4-20). (CEUR Workshop Proceedings; Vol. 3617). CEUR-WS.org. <https://doi.org/urn:nbn:de:0074-3617-4>

Guillotel-Nothmann, Christophe, De Berardinis, Jacopo, Bottini, Thomas, Cathé, Philippe, Daga, Enrico, Daquino, Marilena, Davy-Rigaux, Achille, Gurrieri, Marco, Van Kranenburg, Peter, Marzi, Eleanora, McDermott, James, Merño Peñuela, Albert, Mulholland, Paul, Scharnhorst, Andrea, & Tripodi, Rocco. (2022). Polifonia D1.2 Roadmap and pilot requirements 2nd version (1.0). Zenodo. <https://doi.org/10.5281/zenodo.7116561>

## Acknowledgement

This work has been enabled by the EC project "Polifonia: a digital harmoniser for musical heritage knowledge" Grant agreement ID: 101004746

<sup>3</sup> <https://polifonia-project.github.io/ecosystem/> Access December 4, 2023  
<https://web.archive.org/web/20231204165657/https://polifonia-project.github.io/ecosystem/>