Deliverable D1.2

Draft Data Management Plan

Grant agreement nr	688382
Project full title	Audio Commons: An Ecosystem for Creative Reuse of Audio Content
Project acronym	AudioCommons
Project duration	36 Months (February 2016 - January 2019)
Work package	WPX
Due date	31 July 2016 (M6)
Submission date	31 July 2016 (M6)
Report availability	Public (X), Confidential ( )
Task leader	Surrey
Authors	Mark Plumbley
Document status	Final

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Table of contents

Executive Summary 3

Background 4

1 Admin Details 5

2 Dataset Information 6

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Executive Summary

This Data Management Plan (DMP) provides an analysis of the main elements of the data management policy that will be used by the project with regard to all the datasets that will be generated by the project. The DMP is not a fixed document, but will evolve during the lifespan of the project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Background

The purpose of this Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the project with regard to all the datasets that will be generated by the project.

The DMP is not a fixed document, but will evolve during the lifespan of the project.

The DMP will address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

The approach to the DMP follows that outlined in the "Guidelines on Data Management in Horizon 2020" (Version 2.1, 15 February 2016).

Data set reference and name: Identifier for the data set to be produced.

Data set description: Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

Standards and metadata: Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.

Data sharing: Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.). In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Archiving and preservation (including storage and backup): Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

1 Admin Details

Project Title: Audio Commons: An Ecosystem for Creative Reuse of Audio Content

Project Number: 688382

Funder: European Commission (Horizon 2020)

Lead Institution: Universitat Pompeu Fabra (UPF)

Project Coordinator: Prof Xavier Serra

Project Data Contact: Sonia Espi, sonia.espi@upf.edu

Project Description: The democratisation of multimedia content creation has changed the way in which multimedia content is created, shared and (re)used all over the world, yielding significant amounts of user-generated multimedia resources, big part shared under open licenses. At the same time, creative industries need to reduce production costs in order to remain competitive. There is, therefore, an opportunity for creative industries to incorporate such content in their productions, but there is a lack of technologies for easily accessing and incorporating that type content in their creative workflows. In the particular case of sound and music, a huge amount of audio material like sound samples, soundscapes and music pieces, is available and released under Creative Commons licenses, both coming from amateur and professional content creators. We refer to this content as the 'Audio Commons'. However, there exist no practical ways in which Audio Commons can be embedded in the production workflows of the creative industries, and licensing issues are not easily handled across the production chain. As a result, most of this content remains unused in professional environments. The aim of this project is to create an ecosystem of content, technologies and tools to bring the Audio Commons to the creative industries, enabling creation, access, retrieval and reuse of Creative Commons audio content in innovative ways that fit the requirements of the use cases considered (e.g., audiovisual, music and video games production).Furthermore, we tackle rights management challenges derived from the content reuse enabled by the created ecosystem and research about emerging business models that can arise from it. Our project will benefit creative industries by providing new and innovative creativity supporting tools and reducing production costs, and will benefit content creators by offering a channel to expose their works to professional environments and to allow them to (re)licence their content.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

2 Dataset Information

Individual Dataset Information

Data set reference and name

DS 2.1.1: Requirements interviews

Data set description

Notes/transcripts from structured interviews with creative industry content users in Task 2.1: Analysis of the requirements from creative industries.
WP: WP2 / Task: Task 2.1
Responsible: QMUL (& MTG-UPF)

Standards and metadata

Text documents

Data sharing

Anonymized form to be made available as appendix to Deliverable D2.1: Requirements report and use cases.

Archiving and preservation (including storage and backup)

Stored on project document server.
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 2.2.1: Audio Commons Ontology

Data set description

Definition of Audio Commons Ontology, the formal ontology for the Audio Commons Ecosystem. Data form of D2.2: Draft ontology specification and D2.3: Final ontology specification.
WP: WP2 / Task: Task 2.2
Responsible: QMUL

Standards and metadata

OWL Web Ontology Language

Data sharing

Public

Archiving and preservation (including storage and backup)

Stored on project document server (& github)
Estimated final size (Bytes): 10K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 2.3.1: ACE interconnection evaluation results

Data set description

Results of evaluation of technological solutions for the orchestration/interconnection of the different actors in the Audio Commons ecosystem. Supporting data for deliverable D2.5: Service integration technologies.
WP: WP2 / Task: Task 2.3
Responsible: QMUL (& MTG-UPF)

Standards and metadata

Tabular (e.g. CSV)

Data sharing

Public

Archiving and preservation (including storage and backup)

Project document store.
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 2.5.1: ACE Service evaluation results

Data set description

Results of continuous assessment of ontologies, API specification and service orchestration through the lifetime of the project, including API usage statistics.
WP: WP2 / Task: Task 2.5
Responsible: QMUL (& MTG-UPF)

Standards and metadata

Tabular (e.g. CSV)

Data sharing

Public

Archiving and preservation (including storage and backup)

Project document store.
Estimated final size (Bytes): 1M

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 2.6.1: ACE Service

Data set description

Freesound and Jamendo content exposed in the Audio Commons Ecosystem. Not strictly a “dataset”, rather a service providing access to data.
WP: WP2 / Task: Task 2.6
Responsible: MTG-UPF (& Jamendo)

Standards and metadata

Audio Commons Ontology

Data sharing

Available via ACE service API

Archiving and preservation (including storage and backup)

Dynamic service availability, no plans to provide a “snapshot”.
Estimated final size (Bytes): N/A

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 4.2.1: Semantic annotations of musical samples

Data set description

Results of semantically annotating musical properties such as the envelope, the particular note being played in a recording, or the instrument that plays that note. Supporting data for deliverables D4.4, D4.9, D4.10, D4.11
WP: WP4 / Task: Task 4.2
Responsible: MTG-UPF (& QMUL)

Standards and metadata

Annotations will be stored using standard formats such as JSON and YAML, and Semantic Web formats such as RDF/XML and N3, and following the Audio Commons Ontology
definition.

Data sharing

Public: Access via Audio Commons API

Archiving and preservation (including storage and backup)

ACE Server. Annotation size estimate: 10kBytes per file x 500k files = 5 GBytes
Estimated final size (Bytes): 5 GBytes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 4.3.1: Semantic annotations of musical pieces

Data set description

Results of music piece characterisations such as bpm, tonality or structure. The specific selection of audio properties to include in the semantic annotation will depend on the requirements of the Audio Commons Ontology. Supporting data for deliverables D4.4, D4.9, D4.10, D4.11
WP: WP4 / Task: Task 4.3
Responsible: QMUL (& MTG-UPF)

Standards and metadata

Annotations will be stored using standard formats such as JSON and YAML, and Semantic Web formats such as RDF/XML and N3, and following the Audio Commons Ontology
definition.

Data sharing

Public: Access via Audio Commons API

Archiving and preservation (including storage and backup)

ACE Server. Annotation size estimate: 300kBytes per file x 500k files = 150 GBytes
Estimated final size (Bytes): 150 GBytes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 4.4.1: Evaluation results of annotations of musical samples

Data set description

Results of evaluation of automatic methods for the semantic annotation of music samples. Results may include human evaluations via listening tests, if required. Supporting data for deliverables D4.4, D4.10
WP: WP4 / Task: Task 4.4
Responsible: MTG-UPF (& QMUL)

Standards and metadata

Tabular (e.g. CSV)

Data sharing

Statistical analysis: Public. Listening tests: Data collected and stored according to ethics policy and approval; anonymized result data publicly available.

Archiving and preservation (including storage and backup)

Project document server. Personally identifiable data password-protected or stored securely on paper.
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 4.5.1: Evaluation results of annotations of musical pieces

Data set description

Results of evaluation of automatic methods for the semantic annotation of music pieces. Results may include human evaluations via listening tests, if required. Supporting data for deliverables D4.5, D4.11
WP: WP4 / Task: Task 4.5
Responsible: QMUL (& MTG-UPF)

Standards and metadata

Tabular (e.g. CSV)

Data sharing

Stastical analysis: Public. Listening tests: Data collected and stored according to ethics policy and approval; anonymized result data publicly available.

Archiving and preservation (including storage and backup)

Project document server. Personally identifiable data password-protected or stored securely offline (e.g. paper in locked filing cabinet).
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 4.6.1: Evaluation results of musical annotation interface

Data set description

Results of evaluation of interface for manually annotating musical content, in terms of its usability and its expressive power for annotating music samples and music pieces. The evaluation will be carried out with real users and in combination with the evaluation of Task 5.4. Supporting data for deliverable D4.9
WP: WP4 / Task: Task 4.6
Responsible: MTG-UPF

Standards and metadata

Free text and Tabular (e.g. CSV)

Data sharing

Usability data collected and stored according to ethics policy and approval; anonymized result data publicly available.

Archiving and preservation (including storage and backup)

Project document server. Personally identifiable data password-protected or stored securely offline (e.g. paper in locked filing cabinet).
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 4.7.1: Outputs of integrated annotation technology: Musical content

Data set description

Annotations of Freesound and Jamendo content. Success in Task 4.7 will result in at least 70% of Freesound (musical content) and Jamendo content annotated with Audio Commons metadata as defined in the Audio Commons Ontology. This will incorporate datasets DS 4.2.1 and DS 4.3.1.
WP: WP4 / Task: Task 4.7
Responsible: MTG-UPF & Jamendo

Standards and metadata

Annotations will be stored using standard formats such as JSON and YAML, and Semantic Web formats such as RDF/XML and N3, and following the Audio Commons Ontology
definition.

Data sharing

Available via ACE service API

Archiving and preservation (including storage and backup)

ACE Server
Estimated final size (Bytes): 150 GBytes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 5.1.1: Timbral metadata & ontology of timbral descriptors

Data set description

Timbral metadata in existing content from Freesound (and potentially other sources), supplemented with descriptors from verbal elicitation experiments. Analysis will provide an ontology of timbral descriptors. Data will support Deliverable D5.1.
WP: WP5 / Task: Task 5.1
Responsible: Surrey-IoSR (& MTG-UPF)

Standards and metadata

Annotations will be stored using standard formats such as JSON and YAML, and Semantic Web formats such as RDF/XML and N3, and following the Audio Commons Ontology
definition. Analysis will be stored in free text and tabular form (e.g. CSV).

Data sharing

Existing metadata: Public. Results of verbal elicitation: Data collected and stored anonymously according to ethics policy and approval; result data publicly available.

Archiving and preservation (including storage and backup)

Project document server.
Estimated final size (Bytes): 1M

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 5.2.1: Timbral listening tests

Data set description

Results of listening experiments on timbre perception, carried out to inform the specification of required enhancements to existing metrics, and of modelling approaches
for significant timbral attributes not covered by the prototype system.
WP: WP5 / Task: Task 5.2
Responsible: Surrey-IoSR

Standards and metadata

Tabular (e.g. CSV)

Data sharing

Data collected and stored anonymously according to ethics policy and approval: publicly available.

Archiving and preservation (including storage and backup)

Project document server.
Estimated final size (Bytes): 100k

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 5.3.1: Evaluation results of automatic annotation of non-musical content

Data set description

Results of evaluation of automatic methods for the semantic annotation of non-musical content, including listening tests where appropriate. Annotations will be evaluated against the timbral descriptor hierarchy defined in Task 5.1. Supporting data for Deliverables D5.3, D5.7
WP: WP5 / Task: Task 5.3
Responsible: Surrey-CVSSP (& Surrey-IoSR)

Standards and metadata

Tabular (e.g. CSV)

Data sharing

Data collected and stored anonymously according to ethics policy and approval: publicly available.

Archiving and preservation (including storage and backup)

Project document server.
Estimated final size (Bytes): 100k

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 5.4.1: Evaluation results of non-musical annotation interface

Data set description

Results of evaluation of interface for manually annotating non-musical content, in terms of its usability and its expressive power for annotating . The evaluation will be carried out with real users and in combination with the evaluation of Task 4.6. Supporting data for deliverable D5.5.
WP: WP5 / Task: Task 5.4
Responsible: MTG-UPF

Standards and metadata

Tabular (e.g. CSV)

Data sharing

Usability data collected and stored according to ethics policy and approval; anonymized result data publicly available.

Archiving and preservation (including storage and backup)

Project document server. Personally identifiable data password-protected or stored securely offline (e.g. paper in locked filing cabinet).
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 5.5.1: Outputs of integrated annotation technology: Musical content

Data set description

Annotations of Freesound and Jamendo content. Success in Task 5.5 will result in at least 70% of Freesound (non-musical) content annotated with Audio Commons metadata as defined in the Audio Commons Ontology. This will incorporate datasets DS 4.2.1 and DS 4.3.1.
WP: WP5 / Task: Task 5.5
Responsible: MTG-UPF

Standards and metadata

Annotations will be stored using standard formats such as JSON and YAML, and Semantic Web formats such as RDF/XML and N3, and following the Audio Commons Ontology definition.

Data sharing

Available via ACE service API

Archiving and preservation (including storage and backup)

ACE Server. Annotation size estimate: 100kBytes per file x 200k files = 20 GBytes
Estimated final size (Bytes): 20 GBytes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 6.4.1: Evaluation results of ACE for Creativity Support

Data set description

Results of holistic evaluation of the ACE in the context of Creativity Support. This will include the results of novel methods to assess how the ACE system and tools facilitate creative flow, discovery, innovation and other relevant dimensions of creative work. Supporting data for Deliverables 6.8, 6.12.
WP: WP6 / Task: Task 6.4
Responsible: QMUL (with Industrial Partners)

Standards and metadata

Free text and Tabular (e.g. CSV)

Data sharing

Usability data collected and stored according to ethics policy and approval; anonymized result data publicly available.

Archiving and preservation (including storage and backup)

Project document server. Personally identifiable data password-protected or stored securely offline (e.g. paper in locked filing cabinet).
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 6.5.1: Evaluation results of ACE in music production

Data set description

Results of evaluation of ACE in music production, measure the utilities of ACE in typical music production workflows. The results will include usability data from beta testers available from Waves and students of Queen Mary's Media and Arts Technology (MAT) programme. Supporting data for Deliverable 6.4.
WP: WP6 / Task: Task 6.5
Responsible: QMUL (with Waves)

Standards and metadata

Free text and Tabular (e.g. CSV)

Data sharing

Usability data collected and stored according to ethics policy and approval; anonymized result data publicly available.

Archiving and preservation (including storage and backup)

Project document server. Personally identifiable data password-protected or stored securely offline (e.g. paper in locked filing cabinet).
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 6.6.1: Evaluation results of search and retrieval interfaces for accessing music pieces

Data set description

Results of evaluation of search and retrieval interfaces for accessing Audio Commons music pieces. The data will support assessment of how ACE supports information seeking activities in creative music production using the web-based interfaces created in Task 6.6. Supporting data for Deliverable D6.5.
WP: WP6 / Task: Task 6.6
Responsible: QMUL (with Jamendo)

Standards and metadata

Free text and Tabular (e.g. CSV)

Data sharing

Usability data collected and stored according to ethics policy and approval; anonymized result data publicly available.

Archiving and preservation (including storage and backup)

Project document server. Personally identifiable data password-protected or stored securely offline (e.g. paper in locked filing cabinet).
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 6.7.1: Evaluation results of ACE in sound design and AV production

Data set description

Results of evaluation of ACE in sound design and audiovisual production. The results will include usability data from beta testers available from AudioGaming and students from Surrey's Film and Video Production Engineering BA (Hons). Supporting data for Deliverable D6.6.
WP: WP6 / Task: Task 6.7
Responsible: QMUL (with AudioGaming)

Standards and metadata

Free text and Tabular (e.g. CSV)

Data sharing

Usability data collected and stored according to ethics policy and approval; anonymized result data publicly available.

Archiving and preservation (including storage and backup)

Project document server. Personally identifiable data password-protected or stored securely offline (e.g. paper in locked filing cabinet).
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 7.1.1: Website statistics

Data set description

Website visitor data and alignment with associated project events. Success in Task 7.1 will yield 50 daily unique visitors to the AudioCommons web portal, (excluding bots), increased by at least 50% during time periods influenced by AudioCommons events.
WP: WP7 / Task: Task 7.1
Responsible: MTG-UPF

Standards and metadata

Tabular (e.g. CSV)

Data sharing

Public (following removal of any personally identifiable information)

Archiving and preservation (including storage and backup)

Web server, backed up on project document server. Storage estimate: 1k / visit x 100 visits/day x 300 days = 30MBytes
Estimated final size (Bytes): 30 MBytes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Individual Dataset Information

Data set reference and name

DS 7.5.1: List of Key Actors in the creative community

Data set description

A list of Key Actors in the creative community will be built and maintained to facilitate dissemination activities in Task 7.5. This includes personally identifiable information such as contact details and interests, and will be maintained according to data protection policies.
WP: WP7 / Task: Task 7.5
Responsible: MTG-UPF

Standards and metadata

Text document

Data sharing

Project partners only.

Archiving and preservation (including storage and backup)

Stored on project document server, in compliance with data protection policies.
Estimated final size (Bytes): 100K

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382