D3.1

Data sharing infrastructure, neural content description, rights management and monetisation v1

Project Title	XR mEdia eCOsystem
Contract No.	101070250
Instrument	Innovation Action
Thematic Priority	HORIZON-CL4-2021-HUMAN-01-06
Start of Project	1 September 2022
Duration	36 months

Deliverable title	Data sharing infrastructure, neural content description, rights management and monetisation v1
Deliverable number	D3.1
Deliverable version	V1.0
Previous version(s)	n/a
Contractual Date of delivery	31.12.2023
Actual Date of delivery	09.02.2024
Nature of deliverable	Report
Dissemination level	Public
Partner Responsible	JRS
Author(s)	Werner Bailer (JRS), Maria Pegia (CERTH), Stamatis Samaras (CERTH), Sotiris Diplaris (CERTH), Antonis Karakottas (CERTH), Helmut Neuschmied (JRS), Cláudia Marinho (MOG) Antonio Calvo, Rubén Ramiro (Eviden), Maurizio Montagnuolo (RAI), Theodoros Chiou (IPR), Leander Stähler (KUL)
Reviewer(s)	Antonis Karakottas (CERTH), Nico Heise (DW)
EC Project Officer	Andreea Popescu

Abstract	This report provides the first version of neural cross-modal descriptors, and technology for extraction, matching and indexing, as well as a basic setup of the data sharing infrastructure, including data model definition. This deliverable also provides the concept of the rights management and licensing framework, and an initial version of licensing components.

Keywords	Repository, search, retrieval, content description, data spaces, rights management, licensing, smart contract blockchain

Copyright

Revision History

VERSION	DATE	MODIFIED BY	COMMENTS
V0.1	28/07/2023	Werner Bailer (JRS)	First table of contents
V0.2	23/08/2023	Werner Bailer (JRS)	Updated table of contents
V0.3	27/11/2023	JRS, CERTH	Description of methods, updated dataset list
V0.4	11/12/2023	FINC; MOC, UNIBAS, Atos, RAI, KUL, IPRights	Complete draft
V0.5	11/12/2023	JRS	Version for internal review
V0.6	19/12/2023	CERTH, JRS	Revision after internal review
V0.7	31/01/2024	KUL, IPR	Revision of legal analysis section
V0.8	05/02/2024	DW, JRS	Review and corrections of revised text
V1.0	09/02/2024	DW	Final Document

Glossary

ABBREVIATION	MEANING
A	Affinity matrix
$A_{\{i,j\}}$	Affinity probabilities
API	Application Programming Interface
CSQ	Central Similarity Quantization
CERTH	Centre for Research & Technology Hellas
CMCL	Cross-Modal Centre Loss
d_c	Hash code length
DGCNN	Dynamic Graphic Convolutional Neural Network
DOCH	Discrete Online Cross-modal Hashing
dB	Database
ELK	Elastic Stack
FCMH	Fast Cross-Modal Hashing
Fscore@k	Fscore at k
GDPR	General Data Protection Regulation
GSPH	Generalized Semantic Preserving Hashing
HCOH	Hadamard Codebook based Online Hashing
HMD	Head-mounted Display
JWT	JSON Web Tokens
KDLFH	Kernel-based Discrete Latent Factor Hashing
l_2	Euclidean distance
LAH	Label-Attended Hashing
M	Set of all existing modalities
mAP	Mean Average Precision
MeshCNN	Mesh Convolutional Neural Network
MTFH	Matrix Tri-Factorization Hashing
P	Semantic space
$p_{\{i,j\}}$	Semantic probabilities
PointNet	Point cloud Network
precision@k	Precision at k
RAI	Radiotelevisione Italiana
recall@k	Recall at k
SSAH	Self-Supervised Adversarial Hashing Network
SSL	Secure Sockets Layer
WP	Work Package

Table of Contents

Index of Figures

Index of Tables

1 Executive Summary

This document provides an overview of the results achieved in WP3 within the first 16 month of the project. The delivery date of this deliverable marks a milestone where a first set of services and components from WP3 are released and available for a first integration in the pilots. In addition to these first versions of integratable components, this deliverable also describes earlier work, where first research results are available, and which will be fed into updated or additional services in the second half of the project.

The **repository infrastructure** forms the backbone of the work around data sharing, search and licensing. This document introduces the Neural Media Repository (NMR), which serves as the backend component for indexing and search. Upon ingest of content, it invokes the content description services also presented in this document. In order to enable not only search in the local content repository, but also metasearch across other content sources, connectors are provided. These connectors federate queries to other content sources. This enables querying these content sources with the metadata provided by the sources, though not the full range of descriptors extracted when content is ingested into the NMR. A number of connectors have been implemented, and the connectors integrating the RAI media asset management system and user generated content are described in more detail. There is also a connector to an NMR instance, in order to enable search across multiple instances.

The **content description services** include both readily available analysis services that have been integrated (such as those already available in the vitrivr backend) as well as research on new analysis methods and descriptors. This includes research on training landmark classifiers from weakly annotated training data (i.e. on item level rather than on actual occurrence of the landmark), as well as incrementally adding classes from few samples. In order to support multiple modalities of content, both the application of learned compact descriptors to NeRF renderings have been studied, as well as comprehensive research on training multimodal descriptors has been performed. The use of supervised hashing methods in order to enable multimodal and cross-model retrieval has been studied and comprehensive experiments comparing with mesh retrieval methods have been performed.

User and workflow management describes the initial work on infrastructure for the authentication of users. These components support both the other WP3 services as well as the orchestrator described in D4.1.

Search services describe the components for local search in the NMR of an XReco deployment. The section also describes the metasearch component, which enables federated search across a set of connectors, and the reranking of the obtained search results. It also describes a 3D search tool that has been implemented for testing the underlying technology. Eventually, all search functionality will be integrated in a single user interface (described in D4.1), which enables local search and metasearch, as well as search in the content set offered on the marketplace.

The section on **legal requirements for rights management** offers an overview of general principles and relevant applicable EU copyright law acquis in the context of XReco objectives and operation, which is currently a highly dynamic topic especially with regards to AI-generated output. It further provides a conceptual description of rights management under the light of copyright and related rights. Moreover, it describes the work that has been done in analysing the XReco technologies as well as partners' workflows to identify the copyright-sensitive acts that are relevant for the legal assessment and to provide a first description of licensing components and their possible substitution by means of exceptions and limitations application. Questions related with the qualification of an AI-generated output as copyrighted work or as derivative creation are also discussed. As basis for the technical work it also provides a legal analysis of smart contracts.

The section on **Rights management and licensing tools** describes the technical work on rights management and the microservices that have been developed for this purpose. This includes a rights management service providing the main entry point and orchestrating the other services, which are the smart legal contracts (SLC) engine for creating, validating and executing smart contracts, the blockchain service provider (used for notarising the contracts) and the monetization manager. This latter component is used to feed data on the relevance of input assets for the creation of a new asset into the contract. This topic is strongly related to the reconstruction methods described in D4.1. A survey on data valuation methods, that are able to assess the contribution of source data to a new asset has been performed. For AI-based methods, this topic is closely related to explainable AI (XAI) methods.

The document is complemented by technical documentation (sequence diagrams, API specifications) and a description of datasets used or created by WP3 work.

2 Introduction

This document provides an overview of the results achieved in WP3 within the first 16 month of the project. The delivery date of this deliverable marks a milestone where a first set of services and components from WP3 are released and available for a first integration in the pilots. In addition to these first versions of integratable components, this deliverable also describes earlier work, where first research results are available, and which will be fed into updated or additional services in the second half of the project.

Covering the diverse and multidisciplinary tasks in WP3, this document is structured as follows.

The **repository infrastructure** (Section 3) forms the backbone of the work around data sharing, search and licensing. The section describes the data repository and search backend, and introduces the concept of connectors that allow accessing other repositories or open data sources. The already implemented connectors are described in more detail.

The **content description services** (Section 4) include the underlying research on algorithms for extracting structured metadata and feature descriptions from the different types of content supported by the XReco platform.

User and workflow management (Section 5) describes the initial work on infrastructure for the authentication of users. These components support both the other WP3 services as well as the orchestrator described in D4.1

Search services (Section 6) describes the components for search in the local repository of an XReco deployment as well as for the search across all connected data sources. It also describes a 3D search tool that has been implemented for testing the underlying technology, before it will be fully integrated with the other search components.

Legal requirements for rights management (Section **Fehler! Verweisquelle konnte nicht gefunden werden.**) describes the work that has been done in analysing the XReco technologies as well as partners' workflows to identify the actions relevant for the legal assessment. The section also analyses the relevant European legislation, which is currently highly dynamic topic. As basis for the technical work it also provides a legal analysis of smart contracts.

Rights management and licensing tools (Section 7) describes the technical work on rights management, in particular the smart legal contracts backend and the blockchain infrastructure, as well as the chosen architectural approach. It also describes the analysis of approaches for assessing the impact of source content for automating data valuation, a topic is strongly related to the 3D reconstruction methods described in D4.1.

The annexes of the document provide sequence diagrams defining the intended use of WP3 components, API specifications, as well as a description of datasets used or created in WP3 work.

Finally, and outlook on the work towards the second iteration of this deliverable is provided.

3 Repository Infrastructure

3.1 Overview (JRS)

This section introduces the components establishing the backend of the repository infrastructure. As described in D2.1, an instance of the XReco platform provides a Neural Media Repository (NMR). The NMR provides the indexing and search backend functionality for the content managed by the XReco instance. Note that such an instance of the XReco platform may be a dedicated one for a large organisation (such as a broadcaster) as well as hosted one serving multiple users, with access rights to different subsets of the content. The NMR provides functionality for ingest, which invokes the content analysis services described in Section 4 for extracting descriptors to be indexed. This enables the local search functionality described in Section 6.

In addition to searching the NMR of the local XReco instance, the repository infrastructure provides connectors to enable federated search across other content sources. These connected content sources may be other instances of the XReco platform, APIs of archives, content feeds or content marketplaces, or public content repositories (e.g. Wikimedia). This section describes the approach taken to designing these connectors, and describes specific types of connectors implemented so far. These connectors enable the metasearch functionality described in Section 6.

3.2 Data model (Atos)

From the beginning of the project, we defined an interactive data model using the DB diagram[1] online tool. As we are still in the early phase of integration, this data model has not undergone many changes. For this reason, we will simply redirect to the D2.1 section 3.2.3.3 in which it was first presented and explained in detail.

3.3 Neural Media Repository (UNIBAS)

The *Neural Media Repository (NMR)* is responsible for g*enerating and storing feature descriptors* for individual assets and providing *search functionality* based on these descriptors. It is therefore a cornerstone of XReco's search infrastructure with interfaces to the *Orchestrator (D4.1)* and the *XR Marketplace* through the *Connector infrastructure* (see Section 3.4). In turn, the NMR relies on various extraction services provided by different XReco partners (see Section 4). Its high-level architecture and place within XReco are illustrated in Figure 3-1.

At its core, the NMR supports two major workflows: *data ingest and querying*. Data ingest deals with assets (e.g., a video file or a 3D model) that have not been registered with the NMR and it involves registering those assets and processing them, so that they can be found at a later stage. Querying deals with searching and exploring the assets stored in the NMR for the purpose of retrieving items of interest that can be used in the later content-creation stage. Both workflows are exposed through RESTful APIs and can be used by other services in the XReco system. We use Open API[2] for specification of this service endpoint, in order to simplify the process of integrating them into the XReco system.

The NMR backend is partially powered by pre-existing system components from the *vitrivr* project, which has been developed by UNIBAS in recent years[3].

1 https://dbdiagram.io/home/
[2] See https://www.openapis.org/
[3] See https://vitrivr.org/

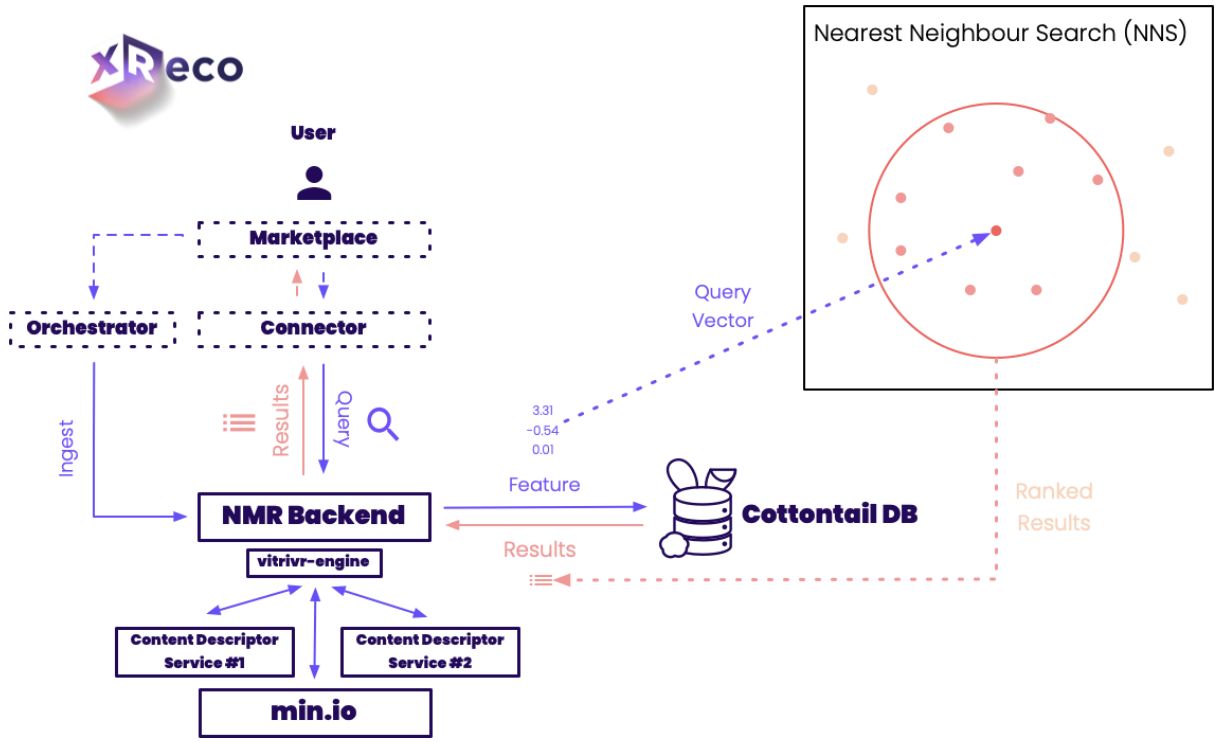


Figure 3-1 NMR backend architecture and its embedding into XReco.

The NMR backend is used by the Connector and the Orchestrator components. It leverages vitrivr-engine and Cottontail DB to facilitate data ingest of assets and search on top of generated descriptors.

3.3.1 vitrivr

Parts of the XReco NMR backend is powered by vitrivr (Rossetto, 2016) – an open-source multi-modal multimedia retrieval stack. vitrivr has been around for several years now and consists of different systems and components, that can also be used independently. Most importantly, vitrivr involves:

- A database for storing and searching *content descriptors (features)* and *metadata*
- A feature extraction and query orchestration engine
- A selection of user-interfaces for querying (web-based and in VR)
- Tools that can be used in the context of multimedia retrieval (e.g., feature extraction services)

In the context of the NMR backend, we rely on the following main components:

vitrivr-engine (Gasser 2024) is the second iteration of vitrivr's feature extraction and query orchestration engine. It supports the two aforementioned workflows: Processing incoming assets in the context of *data ingest* and the orchestration required to execute and fulfil a complex query. vitrivr-engine is a rewrite of the original engine *Cineast.* It has been rebuilt this year for the purpose of being able to support a wider range of applications -- such as XReco -- in which vitrivr is merely a component in an existing ecosystem rather than a monolith. Consequently, vitrivr-engine has a more modular architecture and a more open data model. Furthermore, it features a query specification language to express the wide range of (multimedia) queries.

Cottontail DB (Gasser, 2020) is a database for multimedia retrieval that has been written in Kotlin. Technically, Cottontail DB is a column store. In addition to classical database functionality, Cottontail DB can also store vector representations, often used for describing multimedia data (so called *features* or *content descriptors*). Those

descriptors can then be used to find similar items *through nearest neighbour search*. This search mode can be combined with traditional Boolean queries on (scalar) metadata. In addition to the mere storage capability, Cottontail DB also comes with specialised index structures that can speed-up similarity search. Some of the indexing techniques supported include Product Quantization (PQ) (Jegou, 2010), Locality Sensitive Hashing (LSH) and Vector Approximation Files (VAF) (Weber, 1997).

3.3.2 Data Ingest

Data Ingest is triggered by XReco's O*rchestrator (D4.1)* component, typically based on a user request. A use case could be a user who found some asset (e.g., a video) from an external data source, which should be ingested into the local NMR for later use in content production. In such a case, the user can mark this asset for ingest, which will cause the orchestrator to upload it to the *NMR*.

Once the asset has entered the NMR, it will undergo processing. At a high-level, the processing pipeline involves the following steps in order:

Registration (Asset)	The NMR backend is responsible of processing the original asset and storing it at a central location. In this same step, the NMR backend also assigns a *globally unique identifier* to the asset – the *assetId* – which can be used by all XReco services to refer to it. Currently, we use min.io[4] as a BLOB store for all assets and derivatives.
Decoding	To be able to process the uploaded asset, it must be decoded and converted into a native representation that can be handled by the NMR backend components and *vitrivr-engine*. A typical example is a video file, which is decoded into a stream of still images (the frames) and/or audio samples that are processed individually downstream.
Segmentation (Optional)	Segmentation is a process by which elements generated during decoding are bundled together into *segments* for the purpose of processing. Typically, this is done for efficiency to avoid redundant information. In *vitrvir-engine* and by extensions, the NMR backend, every segment forms a retrievable unit. For example, the stream of still images generated from a video (the frames), are segmented into shots, that is, a contiguous sequence of images that can be represented by a single, representative frame due to self-similarity.
Feature Extraction / Content Description	Feature extraction (also referred to as Content Description, see Section 4) is the process by which features or descriptors are derived from the segments generated upstream. In *vitrvir-engine* and by extensions, the NMR backend, *features* are associated with *segments*, which is what makes those segments retrievable. In the NMR backend, there are two ways to extract a feature. Firstly, vitrivr-engine comes with some integrated feature extraction services. Most importantly, however, vitrivr-engine can interface with external *feature extraction services* developed by other partners, that

[4] See https://min.io/

generate *feature descriptors* for *segment data*. One of those service endpoints is also part of the vitrivr ecosystems (see Section 4.1).

For example, the NMR backend can send the generated segments to the l*andmark feature service build by JRS* (see Section 4.2), to identify and obtain landmark labels for every segment.

Storage All the metadata generated during the data ingest must be persisted. This is the final step in the processing pipeline and leads to information about assets, segments and features to be stored in Cottontail DB for late use. Furthermore, the NMR backend is also responsible for generating and storing derivatives for assets, such as *thumbnails* for *segments*.

3.3.3 Querying

Querying enables to find items that have previously been ingested into the NMR. At a high level, the NMR supports three types of queries:

Similarity Search Similarity search enables the user to find entries (results) in the NMR that are similar to another entry (the query). The way this works is that the NMR backend uses the feature descriptor generated for the query object to perform a nearest neighbour search using Cottontail DB. The top K closest entries are then considered similar and thus relevant for the result set.

In principle, this type of *proximity-based queries* can also be used for querying objects that are not contained in the NMR. For example, *vitrivr-engine* supports the derivation of feature descriptors from raw data (e.g., a user-provided image) on-the-fly. Such a transient feature can then also be used for similarity search as well.

Fulltext Search Many of the descriptors generated by external services are simply text (e.g., labels or generated descriptions). These types of features can be searched by fulltext search, which is also provided by **Cottontail DB** through Apache Lucene[5]. We can use this, for example, to search for generated landmarks or video captions generated by an artificial neural network, such as, CLIP.

Boolean Search / Lookups The NMR backend also allows for more simple search operations, that involve classical Boolean search. The most important instance is that of mapping a retrieved segment to the asset that that segment belongs to, which is the unit of information handled at an XReco system level.

The different types of searches can be encoded in a JSON data structure and sent to the *vitrivr-engine* backend RESTful API. The *vitrivr-engine* then orchestrates the necessary queries in **Cottontail DB** and the transformation (e.g., score normalization or fusion) of the results that follows to return the desired result to the caller. The NMR

[5] See https://lucene.apache.org/

backend simply wraps this core-functionality and provides some purpose-built API endpoints for the most common use cases. It also interfaces search with the XReco services for feature extraction.

3.4 Connectors (Atos)

This sections describes the concept of connectors, which has been introduced to make external data sources available, and feed content into the metasearch and make them available for ingest. After describing the general concept that has been used for all connectors, two specific connectors are described in more detail: the connector for the RAI archive, and the connector to MOG's platform for collecting user generated content.

3.4.1 Connector concept

In XR projects, a diverse range of data types are utilized, including images, videos, text, 3D models, and audio. Therefore, accessing various forms of media is essential. The connector plays a role in gathering and organizing information from various sources. This involves retrieving, validating, and standardizing information into a custom model.

The objective is to obtain a range of diverse information from various sources. Currently, there are five connectors utilized in the metasearch context, each stemming from distinct external sources, namely RAI solr API, DW API, UNIBAS API, Wikimedia API, and SketchFab API. To effectively obtain and compile the information from these sources, each connector requires a unique configuration tailored to its respective source.

To achieve this, the initial step involved incorporating the logic into the configuration. Essentially, the configuration of each connector contains all the necessary logic, meaning that the connector's functionality is dictated by a config file (YAML) rather than the code itself. This particularity enhances the possibility of creating and deploying another connector efficiently.

Another feature is gathering all the information or metadata into a model specially designed and built to suit this project and fulfil its needs. That means, we can harmonize familiar metadata, validate their content and save them in a simple way. Therefore, the connector can provide similar services by abstracting the problem that arises when RAI and DW have the same content but differ in how they are accessed, or the problem when they have different content and ways of access, and XReco needs to categorize it in the same group.

The connectors additionally provide the capability to validate all the information that is provided to verify and amend any information that may be erroneous or contain typos. For instance, the connector can discard a URL containing typographical errors, or it is built improperly.

3.4.2 Architecture and Sequence Diagram

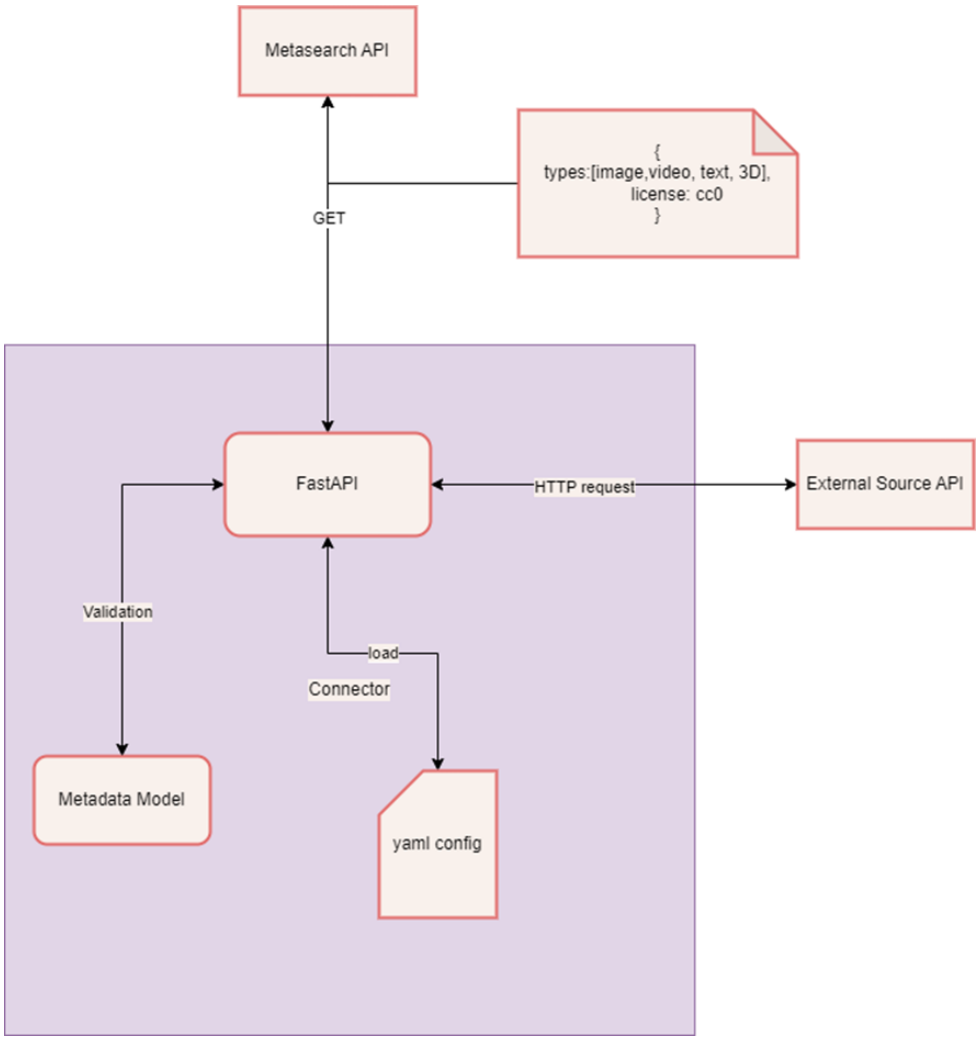


Figure 3-2 Connector Diagram.

Figure 3-2 illustrates the components inside and outside the connector as a micro-service:

Outside of the connector:

- Metasearch API: The interface which communicates with the connector in a double way.
- External Source API: The API where the connectors get data & metadata.

Inside of the connector:

- FastAPI[6]: It is the connector interface with other services. It handles communications with other APIs, such as the external source API or the metasearch API.
- Metadata Model: It is a common schema model built to suit the project's needs.
- YAML config file**:** This file contains the configuration to work correctly with the API that obtains the metadata.

[6] https://fastapi.tiangolo.com/

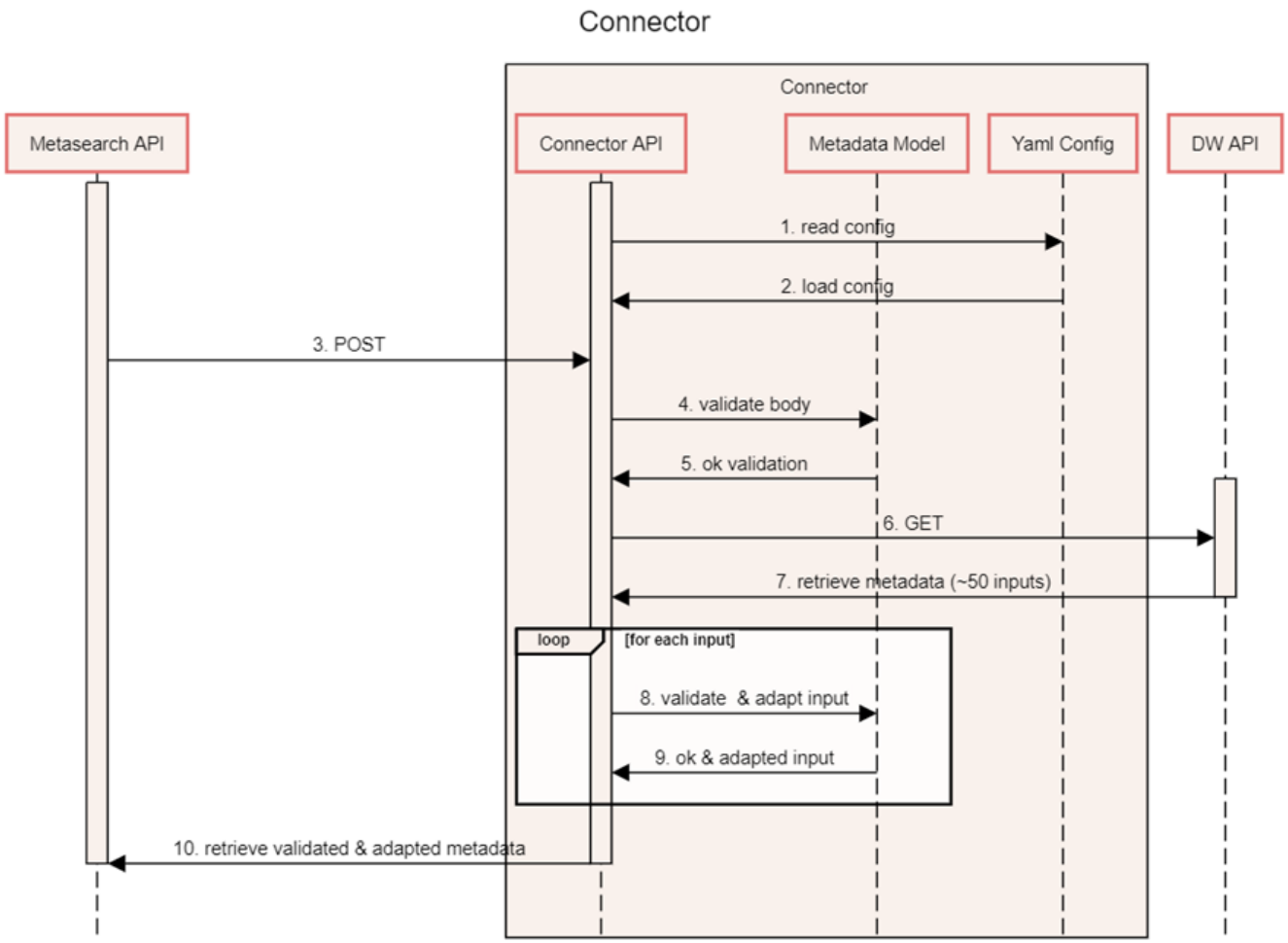


Figure 3-3 Connector Sequence Diagram.

- Within the synchronous communication and micro-service architecture. The sequential steps involved are:
  - (Steps 1 and 2) The connector initiates its instance and loads the configuration written in a YAML file. This configuration changes depending on the source i.e., the company that offers the service to retrieve metadata as RAI, Sketchfab or Deutsche Welle.
  - (Step 3) Metasearch API sends a POST request to the connector based on a term and some filters such as license and type (video, images, text, 3D).
  - (Steps 4 and 5) Connector validates the body and its values using the ad-hoc metadata model.
  - (Steps 6 and 7) The connector will send a GET request to the external source API if the body's validation is proper.
  - (Step 8 and 9) Based on the number of inputs/documents wanted, the connector validates each input/doc, considering the value and format (e.g., conforming URL). If some input is not validated, it is automatically discarded.
  - (Step 10) Finally, the connectors retrieve all the inputs or documents adapted and validated against the metadata model to the metasearch API.

3.4.3 Rai Media Asset Management Connector (Atos & RAI)

The Repository Connector is designed to consume professional content stored in legacy repositories, such as the Rai's archive. Since, for security reasons, the Rai central repository is not accessible from outside the

organisation, we have opted to replicate a selected subset of the internal archive on a cloud platform, accessible from the Internet. The building blocks of this integration are shown in Figure 3-4. The core of the pipeline is the *Worker* (Staging area plus Processing/Adaptation boxes in Figure 3-4), which is responsible for extracting the necessary information from the internal media asset repository, including, video, audio and related metadata, and performing the processing and/or adaptation of the content, such as extracting key frames or generating a proxy version (i.e., low resolution) of the ingested assets that can be used as previews to show the content in a search GUI. Content metadata are stored in a Solr index,[7] which allows for full-text search of the ingested assets. When a user queries the XReco platform through the Metasearch GUI, the request is captured by the Repository Connector and sent to the Solr search engine for further processing. The sequence diagram in Figure 3-5 shows the interaction flow between users, including both internal and external Rai stakeholders, and the system's modules. The workflow is started by a Rai internal user, e.g., a journalist editor, who uploads some content (and related metadata) into the Rai Media Asset Management (MAM) system. Then, the user requests the MAM to export some content to XReco. The MAM instantiates a process workflow that ends with the upload of the content data (i.e., a media file) and metadata (i.e., an accompanying JSON file) from the Rai's private network to the cloud platform. Once the content has been uploaded, the media processor extracts the keyframes from the video, creates a low-resolution version of the video, and parses the JSON file to index the metadata in the Solr server. Finally, the media content and metadata are made available to the XReco users for search and retrieval and subsequent processing.

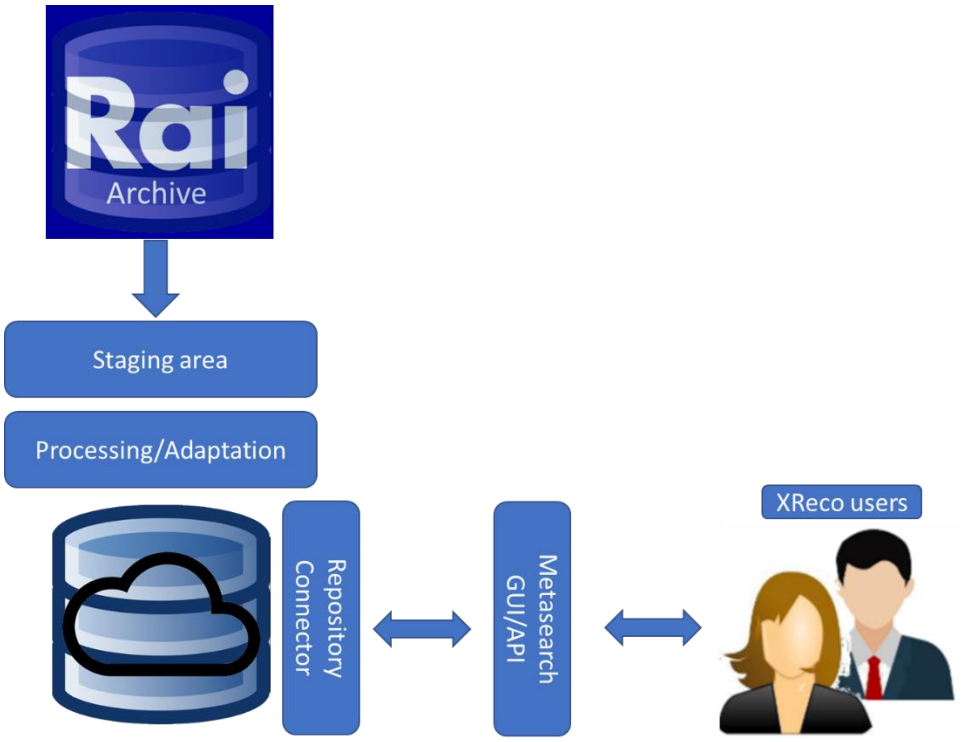


Figure 3-4 Rai Media Asset Management Connector – general workflow

[7] https://solr.apache.org/ (last accessed dec 2023)

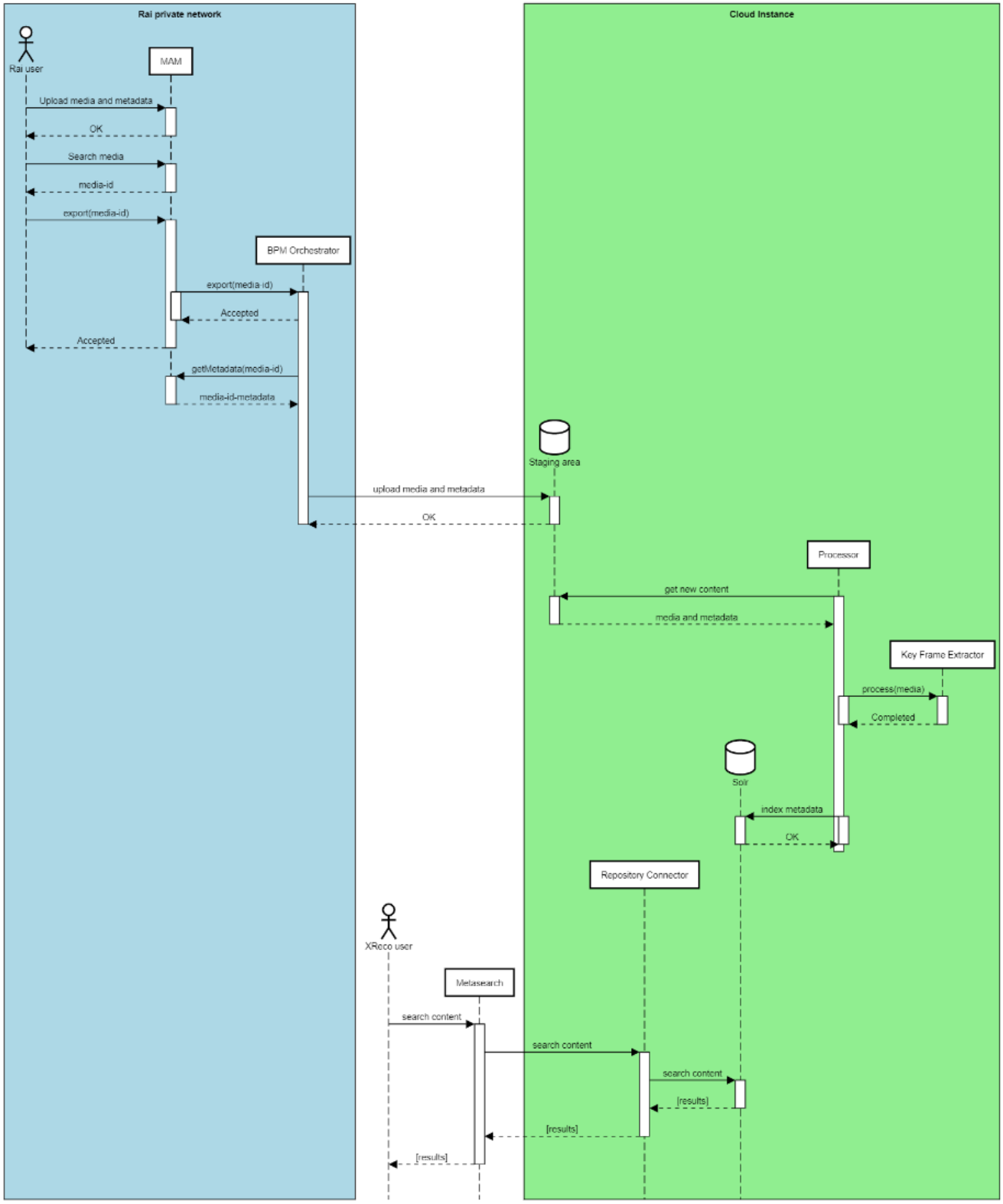


Figure 3-5 Interaction and data exchange between users and the Rai Media Asset Management Connector

3.4.4 User-generated content connector (MOG)

The Repository Connector is designed to consume user-generated content originated from the Crowd Journalism platform, which was developed by MOG Technologies. This integration is visualized in the workflow diagram presented in Figure 3-6.

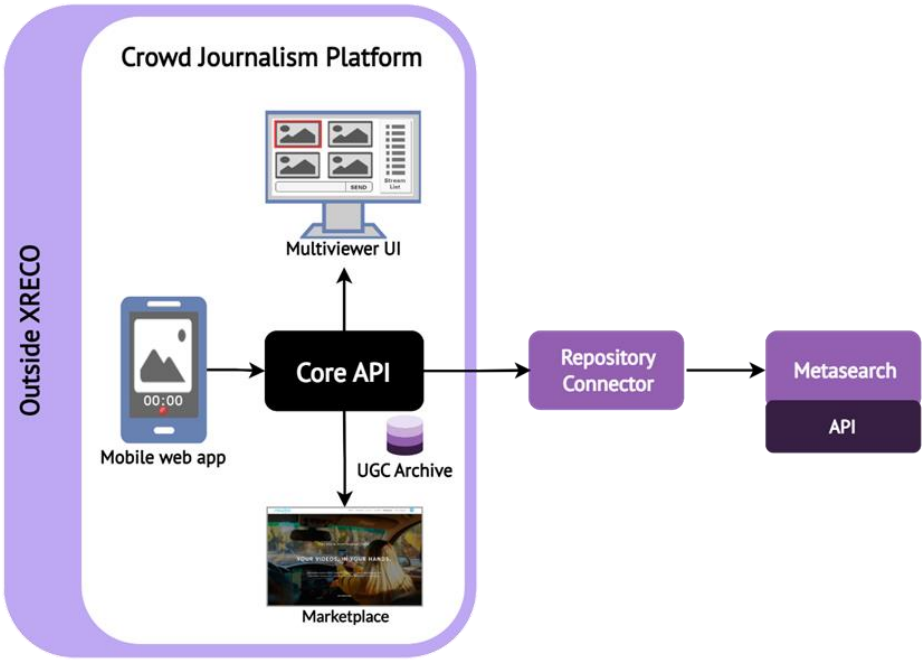


Figure 3-6 User-generated content connector – general workflow.

The Crowd Journalism platform serves as a collaborative space for media production, featuring live news video capture, streaming, and editing capabilities. This empowers both independent journalists and news broadcasting industries to collaboratively generate content by aggregating crowd-sourced news through public participation. On a broader scale, the Crowd Journalism platform consists of five key components:

- **Mobile Web App:** Tailored for smartphones, this application can acquire high-quality video during events and stream it to the Crowd Journalism platform.
- **Multiviewer UI:** Capable of receiving multiple streams from the Crowd Journalism platform, this interface enables real-time selection and watching of news videos. Selected feeds can be published in real-time to external broadcast channels, such as YouTube Live.
- **Core API:** Responsible for processing user inputs, this API ensures the seamless execution of the entire system workflow. It provides an API endpoint that returns the data of videos stored in the User-Generated Content (UCG) Archive.
- **Marketplace:** Allows users to browse and preview contributions from other platform users in a Video on Demand (VOD) format. The Marketplace also facilitates transactions of user-generated content (UGC), enabling users to buy and sell videos.
- **User-Generated Content (UCG) Archive:** This repository stores all recorded videos.

Within the scope of the XReco project, the Crowd Journalism platform will empower end-users to capture videos using the Mobile Web App provided by the platform. This democratizes content creation, allowing citizens at large to contribute their own content both to the Crowd Journalism platform and the XReco platform.

User-generated content has many uses within the XReco project. For instance, if a content creator wants to generate a 3D model of a monument using XReco's reconstruction services but the XReco repository lacks footage from specific angles of said monument, the content creator can leverage the Crowd Journalism platform to capture the missing data. This data will be seamlessly integrated in the XReco repository, contributing to the generation of accurate 3D models of the monument.

The Mobile Web App, showcased in Figure 3-7, can be used with or without user credentials, providing a user-friendly interface to initiate live streams. Accessible through a URL on the web browser, users can effortlessly start a live stream by pressing the "record" button. This interface also includes additional features such as the ability to switch between the front and rear cameras and to mute/unmute audio.

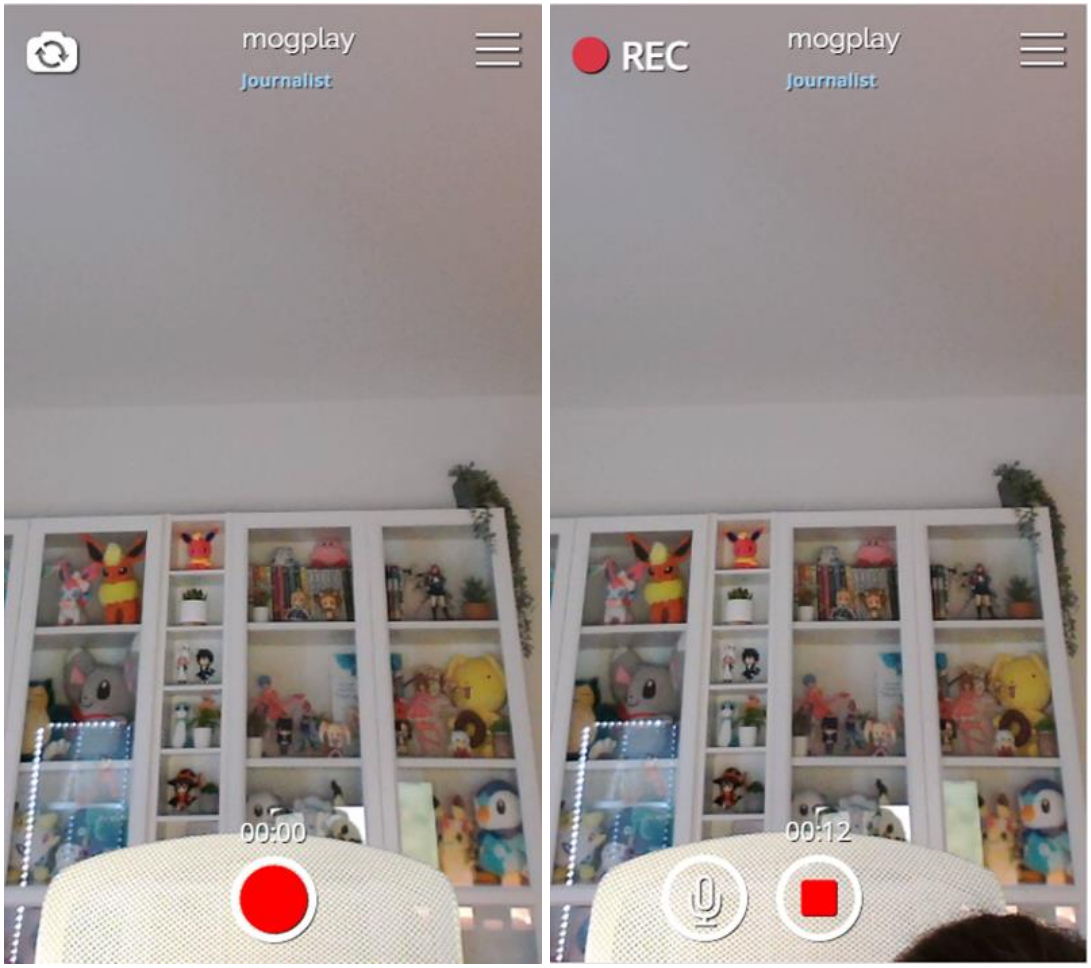


Figure 3-7 User-generated content connector - Mobile web app.

Within the Multiviewer UI, journalists have the capability to watch and edit live streams created in the Mobile Web App in real-time. Once users stop recording, the stream is archived in the User-Generated Content (UGC) Archive, making it accessible on both the Marketplace and the XReco repository.

The information about the archived streams is available on an API endpoint provided by the Core API, enabling the Repository Connector to consume this data. The API Endpoint will provide some essential information about the video, such as the video's creator, URL, thumbnails, and license information.

4 Content description services

4.1 vitrivr analysis services (UNIBAS)

In addition to providing the NMR infrastructure (see Section 3.3), UNIBAS also delivers some content description services to derive general purpose features for similarity search. Mainly, we currently leverage the capabilities of the multi-lingual CLIP model[8]. This model is incorporated into an external Python service, which exposes the feature extraction functionalities through a RESTful API.

The API specification of that service also acts as a template for services developed by other partners with the ultimate aim to have a standardised interface for all the different content description services in XReco. It is also worth mentioning, that *vitrivr-engine* – which forms the foundation for the NMR backend – offers support for such externalised content description services (see Section 3.3.1).

4.1.1 CLIP (Contrastive Language-Image Pre-training)

CLIP, or Contrastive Language-Image Pre-training, stands at the forefront of innovative approaches to multimodal learning, bridging the semantic gap between textual and visual representations. Developed by OpenAI, CLIP extends beyond conventional image recognition models by incorporating a deep understanding of textual and visual contexts.

At its core, CLIP is designed to comprehend images and text in a unified framework. Unlike traditional computer vision models that often rely solely on labelled image datasets, CLIP broadens its horizons by jointly pre-training on a vast corpus of images and associated natural language descriptions. This joint pre-training equips CLIP to connect words and phrases with the visual content they describe, fostering a rich, contextual understanding.

The critical innovation driving CLIP's effectiveness is contrastive learning. Instead of relying solely on positive samples for training, CLIP introduces contrastive pairs—instances where the model learns to differentiate between correct and incorrect associations. This enables CLIP to grasp nuanced relationships between images and text, enhancing its capacity for downstream tasks.

One of CLIP's defining features is its versatility. The pre-trained model is not tailored to a specific application but is a powerful tool for many tasks. Whether it is image classification, zero-shot learning, or even generating textual descriptions for images, CLIP exhibits remarkable performance across various domains without requiring task-specific fine-tuning.

4.1.2 Service

We have implemented a user-friendly service accessible through a RESTful interface. The service accepts the min.io URL of an image or a video and an optional start and end time (to denote a segment). Based on this information, the service obtains the asset from *min.io* and derives a 512-dimensional content descriptor vector for the input. These vectors can then be used for similarity search using the Euclidean distance. The CLIP ViT-B/32 xlm roberta base model was trained with the LAION-5B[9]. The service is capable of leveraging GPU acceleration for the neural network-based descriptor components but can also run in CPU mode only.

[8] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *ArXiv*. /abs/2103.00020

[9] https://laion.ai/blog/laion-5b/

4.2 Weakly-supervised landmark classification (JRS)

For many XR use cases, models of buildings or public spaces, commonly referred to as landmarks, are of great importance. Video data is a particularly interesting source for this purpose, as a moving camera shot may provide several views of the landmark of interest known to be taken at the same time under the same conditions, and there are possibly multiple shots from the same recording in a video. However, if the landmarks of interest are at all annotated, the annotation is weak, i.e., metadata exist on the level of a video file, broadcast, or in the best case a story, but do not provide information about the exact temporal location of views of the landmarks in each frame. Despite their limitations, these weakly annotated videos offer a valuable resource, as they can be used to train models that are able to detect the same landmark in entirely unannotated content, replacing the labor-intensive and time-consuming process of manually labeling training samples. However, the success of deep learning models depends on the quality and size of annotated datasets used for training. The video thus contains "noise" from the point of view of this purpose, such as interviews, shots of anchorpersons, close-up interior views etc. To effectively use these data, the first task is to isolate the relevant ranges of frames showing the landmark of interest. Apart from very popular landmarks, the number of samples may still be small, thus requiring a method capable of learning from a small set of samples (few-shot learning).

To mine images for scene reconstruction, the problem to be solved can be defined as follows. Given is a collection of N videos $\mathrm{V} = \{v_1,\ldots,v_N\}$, where each video v_k is composed of a set of M_k frames , a set of P landmarks $\mathrm{L} = \{L_1,\ldots,L_P\}$, and a set of landmarks contained in a video $A_k = \{L_p,L_q,\ldots\}$ (with $|A_k| \ll P$, and likely $A_k = \{\}$ for some videos). The first step is to mine a training set $\mathrm{T}_p = \{f_j^k | L_p \in A_k \wedge \mathrm{vis}(f,L_p) = 1\}$, where the function $\mathrm{vis}(f,L_p)$ returns 1 if frame the landmark L_p is at least partly visible in frame f, and 0 otherwise. Then a landmark classifier can be trained from L to annotate further videos and increase the set of images mined for reconstruction. By training our deep learning models on the training data mined from the weakly annotated video dataset, we aim to develop algorithms capable of recognizing landmarks with high precision, thus being able to collect an image set providing the basis for reliable scene reconstruction.

4.2.1 Related Work

4.2.1.1 Landmark Recognition

Earlier landmark recognition methods were based on the extraction of local image features, often represented as visual words. With the advent of deep learning, convolutional neural networks (CNNs) were introduced to extract features from images, enabling both landmark classification and the use of similarity measures between pairs of images. (Noh et al., 2017) introduced DELF (DEep Local Features), a fusion of classical local feature methods with deep learning techniques. DELF leverages features from CNN layers and integrates an attention module to enhance recognition accuracy. (Boiarov et al., 2019) extended this by utilizing the center loss function to train CNNs, which penalizes the distances between image embeddings and their corresponding class centers. In handling variations due to different viewpoints, they employed hierarchical clustering to compute centroids for each landmark, effectively managing the variability inherent in landmarks. (Razali et al. 2023) propose a lightweight landmark recognition model using a combination of Convolutional Neural Network (CNN) and Linear Discriminant Analysis (LDA) for feature size reduction. They compared different CCNs and showed that the EfficientNet architecture with a CNN classifier outperformed the other models evaluated. (Yang et al., 2022) train two different variants of the ResNet network architecture (ResNeSt269 and Res2Net200_vd) with an increasing image resolution (step by step) to improve the landmark recognition feature vectors obtained from a CNN. They normalize and merge the embedded layer descriptors of the two models above, doubling their size. As a post-processing step, the retrieval results can be significantly improved by re-ranking methods, e.g. by spatial

verification (Perd'och et al., 2009) which is a method that checks the geometric consistency using local descriptors.

4.2.1.2 Learning from Weakly-labeled Data

The challenge of learning from weakly-labeled data has spurred innovative approaches to filter out potential temporal noise in annotations (Song et al., 2020). A different perspective is presented by (Li et al., 2021), who propose transforming potential noise in weakly-labeled videos into valuable supervision signals. This is achieved through the concept of sub-pseudo labels (SPL), where a new set of pseudo-labels is generated, expanding the original weak label space. This creative approach demonstrates a shift from noise filtering to converting noise into useful information, harnessing the power of weakly-labeled data for improved learning.

The rise of the internet and social media has simplified the acquisition of data relevant to specific classification tasks. In cases where supervision is incomplete and only a portion of training data carries labels, harnessing the abundance of online data becomes crucial. The substantial information available from online sources contributes significantly to robust model development by augmenting labeled training data with additional instances from diverse contexts. But also, images from web search engines like Google or DuckDuckGo tend to be biased toward images where a single object is centered with a clean background and a canonical view- point (Mezuman et al., 2012) and depending on the search term, there might also be pictures included that correspond to a completely different context than intended. (Chen et al., 2015) propose a two-stage CNN training approach for leveraging noisy web data. They initially employ simple images to train baseline visual representation using CNN. Subsequently, they adapt this representation to more challenging and realistic images by capitalizing on the inherent structure of the data and category.

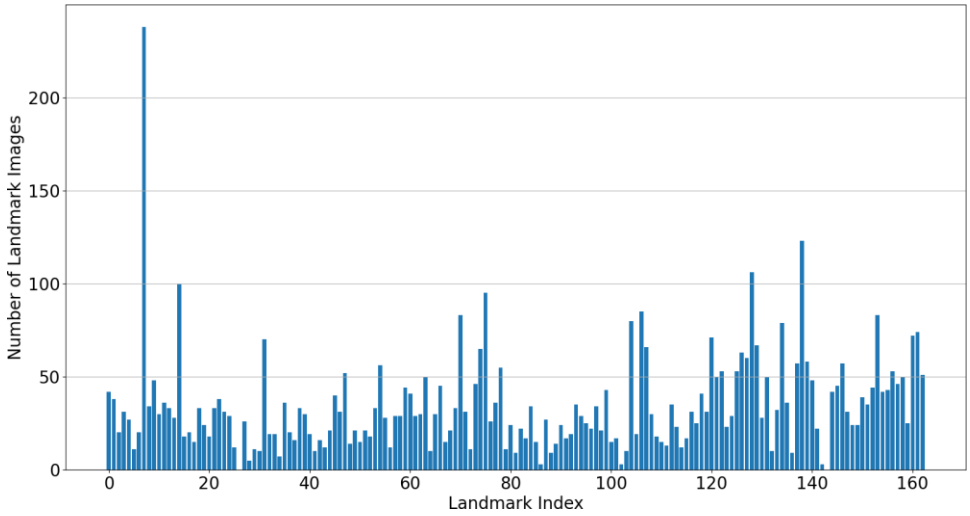
4.2.1.3 Weakly Annotated Landmark Datasets

While there exist many datasets relevant to our task, most of them are fully labeled image datasets. Oxford (Philbin, 2007) and Paris (Philbin, 2008) buildings are early datasets used in landmark recognition and visual place recognition tasks. Over the years datasets grew in size, such as Pittsburg250k (Torii et al., 2013) and data diversity, e.g. including different captures times as in Tokyo 24/7 (Torii et al., 2015). The Google Landmarks datasets v1 (Noh et al., 2017) and v2 (Weyand et al., 2020) have become a widely adopted benchmark for this task. Datasets created for autonomous driving research such as BDD100k (Yu et al., 2018) provide location indexed videos and are thus sometimes used for visual geolocation. However, these datasets show landmarks always from a vehicle point of view, and in contrast to weakly annotated content from media archives the remaining content contains other street views.

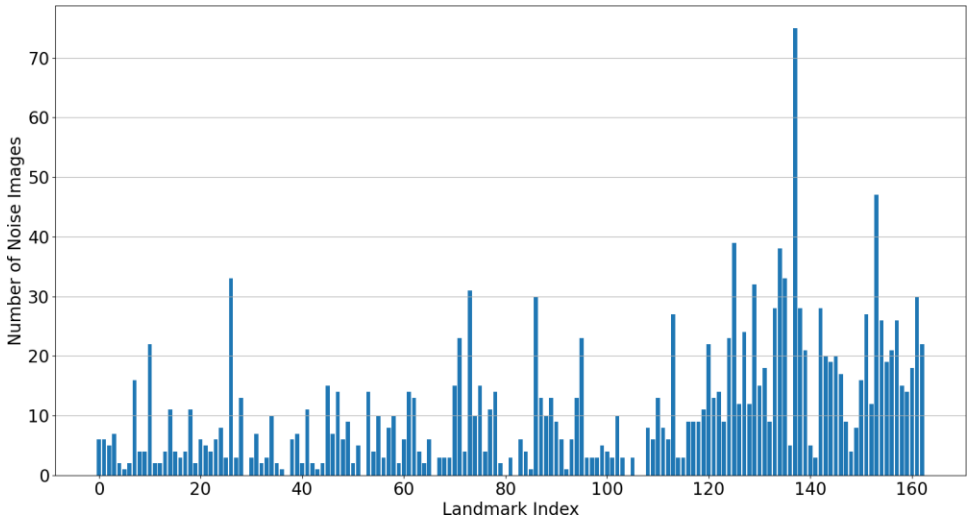
The Italian public broadcaster RAI assembled a dataset with monuments of Italy (Caimotti et al., 2017). The dataset contains about 2,000 clips depicting about 200 monuments from all regions of Italy, mainly acquired from RAI regional newscasts, collected for assessing similarity search in video. Annotations of the monument are provided on clip level. Each clip contains typically a news story, of which one or more shots contain an exterior view of the relevant monument, and in some cases also interior views. Some of the shots may show the monument occluded or in the background (e.g., as backdrop of an interview). In addition, the clips often contain other material of the story, e.g., the anchor in the studio introducing the topic (with an image that shows a view of the monument or something else), views of people in the street, close-up shots of people or interior items etc. As such, the dataset is typical for the type of content and the granularity of annotation to be found in a broadcast archive. The dataset is not available for public download but provided by RAI under a custom license agreement.

V3C (Vimeo Creative Commons Collection) (Rossetto et al., 2019) is a very large dataset (28,450 videos, about 3,800 hours) assembled for benchmarking video retrieval. We have considered amending the existing metadata with landmark annotations for a subset. However, a preliminary experiment found a too small number of clips (using landmark or city names as initial queries).

The lack of a dataset that matches the characteristics of data and annotation granularity found in media archives and that is openly available led to the decision to construct such a dataset by combining commonly used datasets.



(a) Distribution of landmark images in the RAI-MI dataset.



(b) Distribution of noise images for each landmark in the RAI-MI dataset.

Figure 4-1 Statistic information of the RAI-MI dataset.

4.2.2 Weakly Annotated Datasets

We propose the Weakly Annotated Video Landmarks (WAVL) dataset. For evaluating landmark recognition, there is no need to use temporal correlations between successive video frames. This allowed us to simplify our dataset by focusing only on keyframes, which represent either a single frame per video sequence or frames extracted at defined temporal intervals. To mimic a keyframe dataset as extracted from archived video content, with similar characteristics as the RAI monuments of Italy (RAI-MI) dataset, we merged images from two different sources:

the Google Landmarks v2 dataset (Weyand et al., 2020) and the V3C (Rossetto et al., 2019) video dataset. These combined sources allowed us to create a dataset that contains sets of keyframes as they would be extracted from one video, containing both keyframes of a particular landmark as well as unrelated keyframes (noise). The video is annotated with the landmark visible in the subset of keyframes taken from Google Landmarks.

For each set of keyframes representing a video, we combined on average 30 associated images from the Google Landmarks v2 dataset with on average 11 keyframes (noise images) from one of the videos in the V3C1 dataset. This process has been done for 141 landmarks, resulting in a dataset of 5,770 images. The dataset has been made available at https://github.com/XRecoEU/WAVL-Dataset.

To perform a comparative evaluation on the RAI-MI dataset, we selected a subset of 163 different videos of landmarks representing buildings and extracted keyframes from them. The key figures for both datasets are shown in Table 1. The distributions of the landmark and noise images for the RAI-MI dataset vary more strongly, thus we plot them in Figure 3 For several cases the number of samples is in the range of 10 or less, so that the problem can at least partly be considered a few-shot learning problem.

Table 1 Key figures for images per landmark from the RAI-MI and WAVL data sets (LM: landmarks, Std Dev.: standard deviation).

Dataset	LM	LM Images	Noise Images	Mean LG Images	Std Dev. LM Images	Mean Noise Images	Std Dev. Noise Images
RAI-MI	163	5620	1871	34.55	26.92	10.63	10.60
WAVL	141	4230	1592	29.42	1.33	11.29	1.07

4.2.3 Implemented landmark classification service

The implemented landmark classification treats the problem as a fine-grained image classification task. To realize this, we build on the Attentive Pairwise Interaction Network (API-Net) (Zhuang et al., 20020) with EfficientNet B3 (Tan et al., 2019) as the backbone architecture. However, the presence of weak annotations introduces a challenge wherein a substantial portion of training images contains incorrect labels. To address this issue, we have explored two different approaches.

4.2.3.1 Semantic Segmentation Prefiltering

In the first approach (Swin+API-Net) (see Figure 4-2), given our focus on landmarks associated with buildings, we leverage a Swin transformer network for semantic image segmentation (Liu, 2021). This process enables us to filter out images where structures like buildings or walls are observable, providing insights into the size of these areas. Additionally, we utilize information about the recognition of extensive regions where human figures are present, aiding in identifying anchorpersons or interior views with close-up shots, discarding these images as well. All images filtered in this process are put into a *None* class for the training process. This strategy of region-based filtering contributes to balancing the impact of keyframes unrelated to the landmark on the training process.

For inference, a simplified approach is adopted for landmark classification on images, using only the retrained EfficientNet backbone with the API-Net classification layer. This allows a straightforward evaluation of the performance of the trained model on individual test images, focusing solely on the ability of EfficientNet to recognise the trained landmarks.

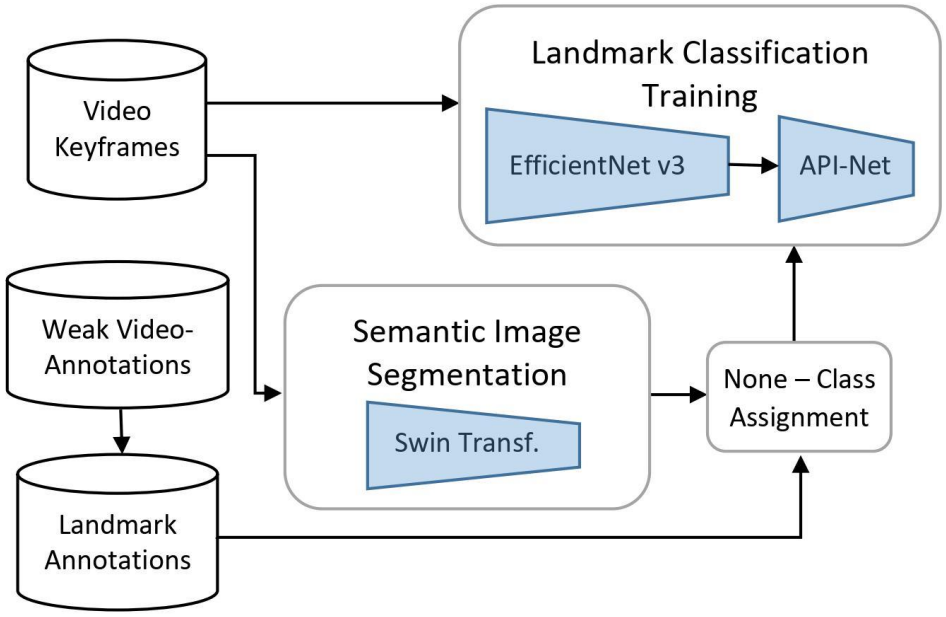


Figure 4-2 Swin+API-Net: Training of the EfficientNet B3 network with API-Net from a filtered set of keyframes.

The corresponding landmark class labels are changed to None class if no or only small building areas are detected in the images.

4.2.3.2 Web Image Mining

In the second approach (CDVA+API-Net) (see Figure 4-3), we perform a targeted web search for relevant landmark training images using the landmark information embedded in the weak video annotations. To facilitate this, we employ the DuckDuckGo image search engine, retrieving a set of 40 images for each landmark term. Recognizing that web data can introduce considerable noise, our selection process is refined. We only retain images that show similarity to keyframes extracted from the weakly annotated videos. This similarity check is done by computing CDVA (Duan et al., 2019) descriptors from the images. We use the learned component of the descriptor, which is binarized to obtain a 512 bit binary vector that can be efficiently matched using Hamming distance. To filter the web images, we compare them to the keyframes and keep only those that have a similarity score above a certain threshold. We use a threshold of 0.6 to select web images in the first step. However, to account for the variability of web images, in a subsequent step we compare the remaining web images with the previously selected ones. If the similarity score exceeds a slightly lower threshold of 0.58, these previously rejected images are also used. Table 2 gives an overview of the web images obtained for the RAI-MI and the WAVL landmarks. Following the filtering of web images, we proceed to train an EfficientNet B3 using the API-Net framework, mirroring the process employed in our first approach. This strategy capitalizes on the wealth of online resources while maintaining a rigorous validation process to ensure the quality and relevance of the acquired images.

Once the training has been completed, the images are again evaluated using only the EfficientNet B3 backbone and the classification layer of API-Net.

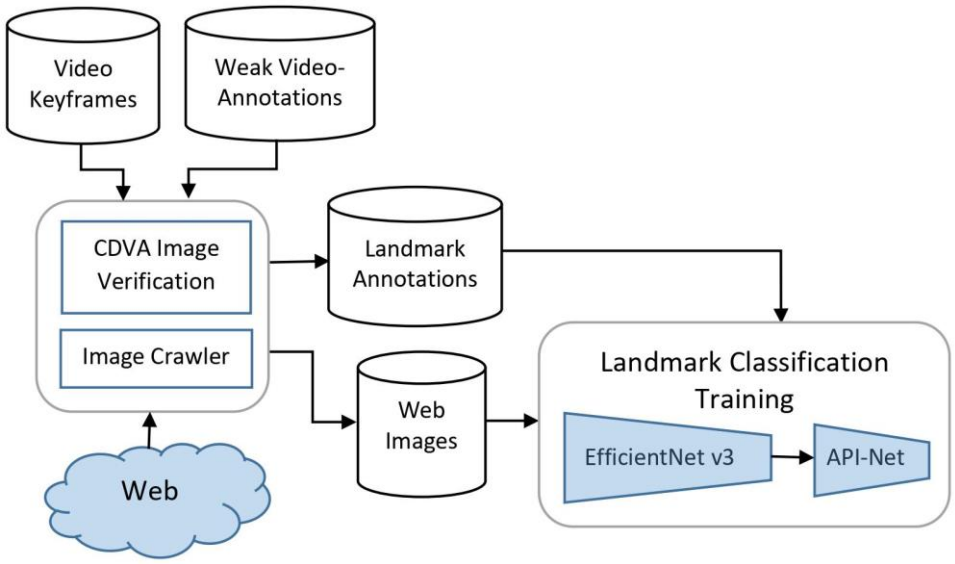


Figure 4-3 CDVA+API-Net: Training of the EfficientNet B3 network with API-Net by using images which are crawled from the web.

Only those web images are used which have sufficient similarity (based on CDVA descriptors) with one of the video keyframes.

Table 2: Images obtained using the web image mining process for the RAI-MI and the WAVL datasets.

Dataset	Landmarks	Downloaded images	Rejected images	Landmarks ≥ 1 image	Landmarks ≥ 10 images
RAI-MI	163	6130	1177	162	153
WAVL	141	5585	166	140	139

4.2.4 Evaluation

We evaluate the two proposed training approaches on both the RAI-MI and WAVL datasets. In the case of the WAVL dataset, we also compare our approach with a state-of-the-art landmark recognition method to establish a baseline for the dataset. Given the composition of the WAVL dataset, which uses Google Landmarks v2 images, we chose to compare our approaches with a well-performing method on this benchmark dataset.

In this context, we considered the solution developed by the smlyaka team (Yokoo et al., 2020), for which the source code is publicly available[10]. It has demonstrated exceptional performance, winning first place in the Google Landmark Retrieval 2019 Challenge and third place in the Google Landmark Recognition 2019 Challenge on Kaggle.

This approach involves the initial pre-training of a ResNet-101 backbone on ImageNet and the Google Landmark Dataset v1 (GLD-v1) (Noa et al., 2017) training dataset. This pre-trained model has not been made available. We thus start from a model pre-trained on ImageNet. To compensate for pre-training on landmarks, we extended the number of training epochs on the WAVL dataset to 14 (instead of 5 as reported by the smlyaka team).

The authors propose an automated data cleaning process to remove wrong annotations. The cleaning process involves a three-step approach that uses spatial verification to filter images by k-NN search. The authors use RANSAC (Fischler et al., 1981) with affine transformation and deep local attentive features (DELF) (Noa et al.,

[10] https://github.com/lyakaap/Landmark2019-1st-and-3rd-Place-Solution

2017) for spatial verification. If the count of verified images reaches a certain threshold, the image is added to the cleaned dataset. We use this approach once without (labelled smlyaka) and once with the data cleaning process (labelled smlyaka with data cleaning).

In evaluating our approaches, we employed a comprehensive suite of evaluation metrics. In addition to precision and recall we use balanced accuracy (BA), the mean of true positive and true negative rate (Brodersen et al., 2010) and symmetric balanced accuracy (SBA) (Gösgens et al.,2021), which aims to eliminate bias by the choice of the positive class. These metrics are related to a specific recognition threshold.

In line with the Google Landmarks benchmark (Weyand et al., 2020), we also use Global Average Precision (GAP, originally proposed as micro-AP in (Perronnin et al., 2009)) as a metric. GAP differs from the more commonly used mean average precision (MAP) in that the mean of precision at rank for all relevant returned results is determined, not taking the number of ground truth positives into account.

Evaluation data are selected from videos which are not used for training data.

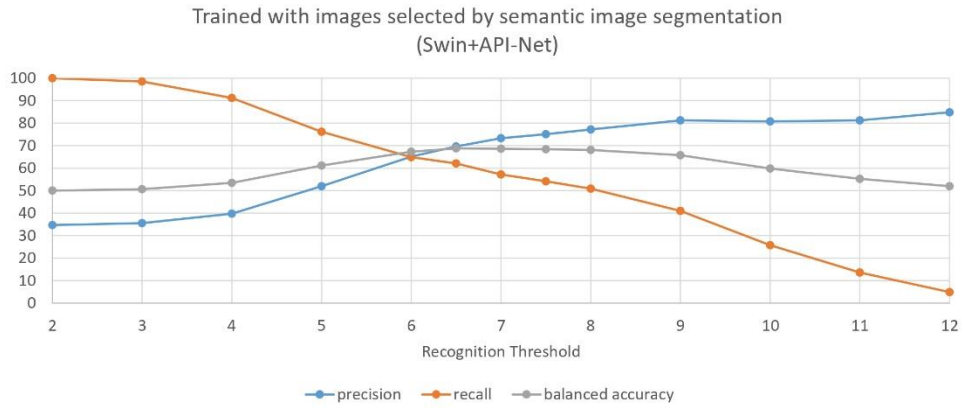


Figure 4-4 Threshold dependent results on the WAVL dataset for the approach Swin+API-Net.

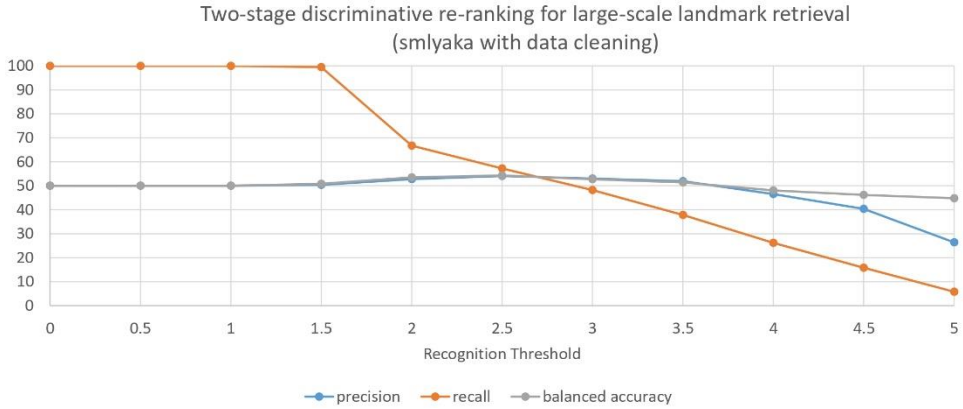


Figure 4-5 Threshold dependent results on the WAVL dataset for the approach smlyaka with data cleaning.

Table 3 Evaluation results for the RAI-MI and WAVL dataset.

Method	GAP	Threshold	Precision	Recall	BA	SBA
API-Net	54.67	7.5	86.80	41.53	**63.67**	**63.41**
Swin+API-Net	**56.09**	7.5	84.88	43.54	63.39	63.05
CDVA+API-Net	44.50	7.5	88.26	27.43	59.69	71.02

(a) Results on the RAI-MI dataset (best score in bold)

Method	GAP	Threshold	Precision	Recall	BA	SBA
smlyaka	47.34	2.0	53.56	**66.31**	54.51	54.64
Smlyaka with data cleaning	47.59	2.5	54.08	57.22	54.31	54.32
API-Net	37.43	9.0	56.27	32.22	53.59	53.99
Swin+API-Net	**53.86**	6.5	69.63	62.06	68.76	68.96
CDVA+API-Net	53.03	6.0	**79.77**	52.57	**70.61**	**72.43**

(b) Evaluation results for the RAI-MI and WAVL dataset.

Table 3 lists the results of the two datasets. The values for precision, recall, BA and SBA are given for a specific threshold, which is determined with respect to a maximum value of BA. The variation of these metrics as a function of the threshold value can be seen in Figure 4-4 and Figure 4-5 for methods Swin+API-Net and smlyaka with data cleaning, respectively.

Filtering out training images using image region classification (Swin+API-Net) has very different effects on the two datasets. In the RAI-MI dataset, the filtered images sometimes contain information that adds value to landmark recognition. These could be, for example, interior views of buildings or studio backgrounds where a newsreader is visible. On the other hand, selecting images by image region classification from the WAVL dataset significantly improves recognition results. In any case, filtering out images reduces the risk of learning unrelated information.

Using web images for training does not provide good results for the RAI-MI dataset. As can be seen in Table 2, one reason is that the web search for this dataset did not yield a sufficient number of usable images for all landmarks. For three of the 163 landmarks, the annotation is incorrect, which means that no landmark related images could be found on the web. Nevertheless, this approach achieves high precision values. We believe that this is due to the similarity constraint in the selection of web images, which will reduce the risk of false positives but also harm the diversity of samples. Another aspect is that in both datasets there are a relatively large number of images for each landmark. If this were not the case, the web-based approach would have the advantage that sufficient images for such landmarks could still be found on the web for training.

The comparison with the winning method of the Google Landmark Retrieval 2019 Challenge (Yokoo et al., 2020) shows that while it performs best in terms of recall, it is otherwise outperformed by one of the proposed approaches on the WAVL dataset. It is particularly interesting that the data cleaning proposed in this approach does not provide any improvement under these conditions. This is because the frequency of similar images is not a good criterion for selecting training images in this use case.

For the purpose of mining images for 3D reconstruction the fact that the proposed approaches have higher precision than other methods is beneficial, as it improves the robustness of the reconstruction process, in particular for methods such as NERF, that do not include a feature matching and filtering step.

4.2.5 Incremental Training

Once the Swin-APInet model has been trained, it should be possible to extend the model with additional landmark classes. This should avoid having to train the classifier from scratch on a huge dataset each time new classes are added. Inspired by (Wang et al., 2020) method for few-shot learning, we tested the following approach:

1. **Model training for the incremental classes:** Training of the Swin-APInet only for the new landmarks by using the backbone network of the original trained model. By doing that the backbone will not be adapted.
2. **Join network:** To obtain a network that can classify both the old and the new landmarks, the two networks are merged. For this purpose, the same backbone network is used again and the classification layers of the two models are connected with each other.
3. **Fine-tuning:** To fine-tune the result network a retraining is done by using the same number of keyframes for each landmark.

The landmark classes trained against a “None” class.

4.2.5.1 Train Dataset

Several test datasets were generated from the RAI dataset, featuring 163 labeled landmarks with a minimum of 9 landmark images per class. The evaluation data was created with care to ensure that no video (from which the keyframes are extracted) used for evaluation was also used for training. Consequently, some landmarks within the evaluation data may present views that were not observed within the training data, such as interior or outdoor views. This generally leads to lower classification rates.

In the first step two lists of evaluation keyframes have been created. Those where one of the 163 landmarks is visible and those where none of these landmarks are present (for the “None” class).

In the following description the initially trained classes are called base classes. We refer to the additionally trained classes as the incremental classes.

For the evaluation different datasets for base and incremental classes have been created (Table 4).

Table 4 Different evaluation datasets, which contain the specified number of image samples for basic and incremental classes.

Base classes		Incremental classes	
Numb. classes	**Numb. images**	**Numb. classes**	**Numb. images**
143	673	20	93
153	724	10	42
158	742	5	24
162	761	1	5
163	766	0	0

To each of these datasets, 309 images are added in which none of the 163 landmarks appear (“None” class).

4.2.5.2 Evaluation

As it is used for evaluation of Swin-APInet itself in addition to precision and recall we use balanced accuracy (BA) and symmetric balanced accuracy (SBA). These metrics are related to a specific recognition threshold. Other metrics listed are Global Average Precision (GAP) and Mean Average Precision (MAP).

We found out that the fine-tuning (retraining) step of the joined network does only improve the results, if also the backbone is retrained. More than 20 training epochs will not give better results. In difference to that, for the training of the models for the base and incremental classes 100 epochs has proven itself. The fine-tuning step is carried out with the same number of image examples for all landmark classes. To achieve this, we determine the minimum number of images of the landmarks available and then this number of examples per landmark is used for training.

Table 5 and Table 6 show the evaluation results of the combined retrained network for different numbers of incrementally trained classes/landmarks. For comparison, the result of the network trained with all landmarks from scratch is also included (number inc. landmarks: 0).

Table 5 Different evaluation datasets which contain the specified number of image samples for basic and incremental classes.

Number original landmarks	Number inc. landmarks	MAP all	GAP all	Threshold	BA all	SBA all
163	0	59.09	50.51	6.5	64.44	63.94
162	1	64.05	51.71	7.5	65.96	66.05
158	5	59.57	51.47	6.5	64.03	63.79
143	10	63.03	51.77	7.0	64.9	64.99
153	20	56.35	48.52	8.0	62.54	63.71

Table 6 Classification results for the retrained joined network.

Number original landmarks	Numb inc. landmarks	MAP all	GAP all	Threshold	BA all	SBA all
163	0	59.09	50.51	7.0	63.94	63.93
162	1	**64.05**	51.71	7.0	**65.74**	**65.53**
158	5	59.57	51.47	7.0	64.02	64.23
143	10	63.03	**51.77**	7.0	64.9	64.99
153	20	56.35	48.52	7.0	62.15	62.04

The evaluation data set is composed of 766 images of all 163 classes and 309 None class images. The BA and SBA values are given for a threshold that yields the best results on average.

4.2.6 Training of incremental classes by using examples of base classes

In the above results, the incremental classes have been trained with images showing these landmarks and the "None" class images, which do not show the incremental and base classes as well.

A variant of this method is to use also images from the base classes for incremental class training. In this case, the base class images are handled as "None" class examples. To restrict the training data, we first find the smallest number of available image examples per class. We then select this number of images for all base classes and label them as the "None" class.

In Table 7 and Evaluation data set is composed of 5 images from the incremental classes and 761 images from the base classes (rated as None class). The value nan (not any number) indicates that the metric could not be calculated (division by zero).

Table 8 (for one incremental class) and Table 9 and Table 10 (for 10 incremental classes) you can see the different behavior of these two variants for model training of the incremental classes. It can be observed that the threshold can be chosen higher (there must be higher confidence values for the examples of the incremental classes) when the images of the base class are used for the "None" class. In addition, the discrimination between the incremental classes and the other classes works better.

Table 7 Results from network model trained with one incremental class (with locked backbone).

Threshold	TP	FP	TN	FN	Accuracy novel	Precision novel	Recall novel	BA novel	SBA novel
0	3	132	629	2	82.51	2.22	60	71.33	61.14
1	2	51	710	3	92.95	3.77	40	66.65	59.16
2	1	17	744	4	97.26	5.56	20	58.88	55.70
3	0	0	761	5	99.35	nan	0	50	nan
4	0	0	761	5	99.35	nan	0	50	nan
5	0	0	761	5	99.35	nan	0	50	nan

Evaluation data set is composed of 5 images from the incremental classes and 761 images from the base classes (rated as None class). The value nan (not any number) indicates that the metric could not be calculated (division by zero).

Table 8 Results from the network model trained with one incremental class (with locked backbone), where examples from the base classes have been labelled as None class.

Threshold	TP	FP	TN	FN	Accuracy novel	Precision novel	Recall novel	BA novel	SBA novel
0	4	10	751	1	98.56	28.57	80	89.34	76.78
1	3	3	758	2	99.35	50	60	79.8	77.34
2	3	1	760	2	99.61	75	60	79.93	83.65
3	3	0	761	2	99.74	100	60	80	89.93
4	1	0	761	4	99.48	100	20	60	79.87
5	0	0	761	5	99.35	nan	0	50	nan

Evaluation data set is composed of 5 images from the incremental classes and 761 images from the base. The value nan (not any number) indicates that the metric could not be calculated (division by zero).

*Table 9 Results from network model trained with **10 incremental classes** (with locked backbone).*

Threshold	TP	FP	TN	FN	Accuracy novel	Precision novel	Recall novel	BA novel	SBA novel
0	33	415	312	6	45.04	7.37	84.62	63.77	58.25
1	33	415	312	6	45.04	7.37	84.62	63.77	58.25
2	33	372	355	6	50.65	8.15	84.62	66.72	59.98
3	33	247	479	7	66.84	11.79	82.5	74.24	64.71
4	29	108	617	12	84.33	21.17	70.73	77.92	68.77
5	19	43	681	23	91.38	30.65	45.24	69.65	66.67
6	9	21	703	33	92.95	30	21.43	59.26	61.01
6.5	6	8	716	36	94.26	42.86	14.29	56.59	62.81
7	4	3	721	38	94.65	57.14	9.52	54.55	65.31
7.5	4	3	721	38	94.65	57.14	9.52	54.55	65.31
8	1	2	722	41	94.39	33.33	2.38	51.05	57.52

Evaluation data set is composed of 42 images from the incremental classes and 724 images from the base classes (labeled as None class)

*Table 10 Results from network model trained with **10 incremental classes** (with locked backbone), where examples from the base classes have been labelled as None class.*

Threshold	TP	FP	TN	FN	Accuracy novel	Precision novel	Recall novel	BA novel	SBA novel
0	25	3	721	17	97.39	89.29	59.52	79.55	86.52
1	25	3	721	17	97.39	89.29	59.52	79.55	86.52
2	25	3	721	17	97.39	89.29	59.52	79.55	86.52
3	25	3	721	17	97.39	89.29	59.52	79.55	86.52
4	25	3	721	17	97.39	89.29	59.52	79.55	86.52
5	24	2	722	18	97.39	92.31	57.14	78.43	86.69
6	20	1	723	22	97.00	95.24	47.62	73.74	84.94
6.5	17	1	723	25	96.61	94.44	40.48	70.17	82.86
7	12	1	723	30	95.95	92.31	28.57	64.22	79.19
7.5	10	1	723	32	95.69	90.91	23.81	61.84	77.59
8	6	1	723	36	95.17	85.71	14.29	57.07	73.78

The evaluation data set is composed of 42 images from the incremental classes and 724 images from the base classes.

After fine-tuning the joined network, the evaluation results for the two approaches of the training of the incremental classes can be seen in Table 11 andTable 12.

Table 11 Classification results for the retrained joined network (153 base and 10 incremental classes) in dependence if base classes are used for the incremental training step or not.

Method for inc. training	Numb inc. landmarks	MAP	GAP	Threshold	BA	SBA
without base class examples	10	63.03	51.77	7.0	64.9	64.99
With base class examples	10	64.57	51.96	7.0	64.95	64.99

The evaluation data set is composed of 766 images from base and incremental classes and 309 None class images.

Table 12 Classification results for the retrained joined network. Base classes are used for incremental training.

Number original landmarks	Numb inc. landmarks	MAP all	GAP all	Threshold	BA all	SBA all
163	0	59.09	50.51	7.0	63.94	63.93
162	1	63.79	51.73	7.0	**65.87**	**65.64**
158	5	59.77	51.65	7.0	64.49	64.69
143	10	**64.57**	**51.96**	7.0	64.95	64.99
153	20	58.12	49.09	7.0	63.36	63.27

The evaluation data set is composed of 766 images of all 163 classes and 309 None class images.

4.2.7 Separate evaluation of incremental classes

In the following, we have investigated the influence of incremental training on the recognition of new classes. We compare the results with those obtained by training all classes from scratch (see Figure 4-6).

Just looking at the model trained only for the new classes, it makes a big difference whether images from the base classes are used in this training. If this is the case, then the incremental classes can be recognized with higher threshold values. Therefore, the best classification threshold value for the base class examples (threshold value 7.0) is also suitable for the incremental class examples.

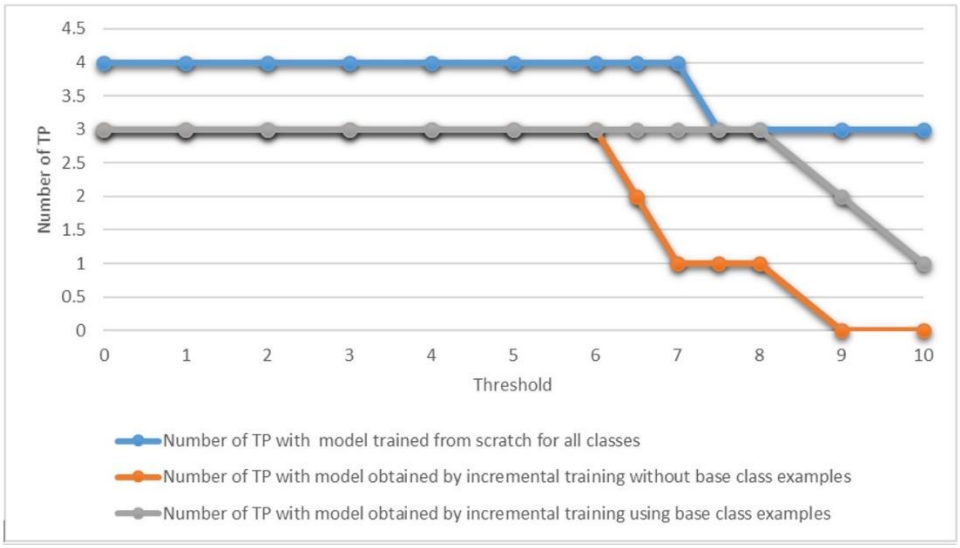


Figure 4-6 Classification results of five images from one incremental class for the cases where the model was trained from scratch with all class examples.

4.2.8 Separate evaluation of base classes

We also evaluated the effects of incremental training on the base classes. For this we use the F_1 metric which is the harmonic mean of precision and recall. The reason for this is that there are no TNs in these tests (no examples of None classes were used) and F_1 is calculated from TP, FP, and FN values:

$$F_1 = \frac{2 \cdot TP}{2 \cdot Tp + FP + FN}$$

As shown in Figure 4-7, incremental training does not have a big impact on the classification results. In fact, we achieve even better results for higher threshold values.

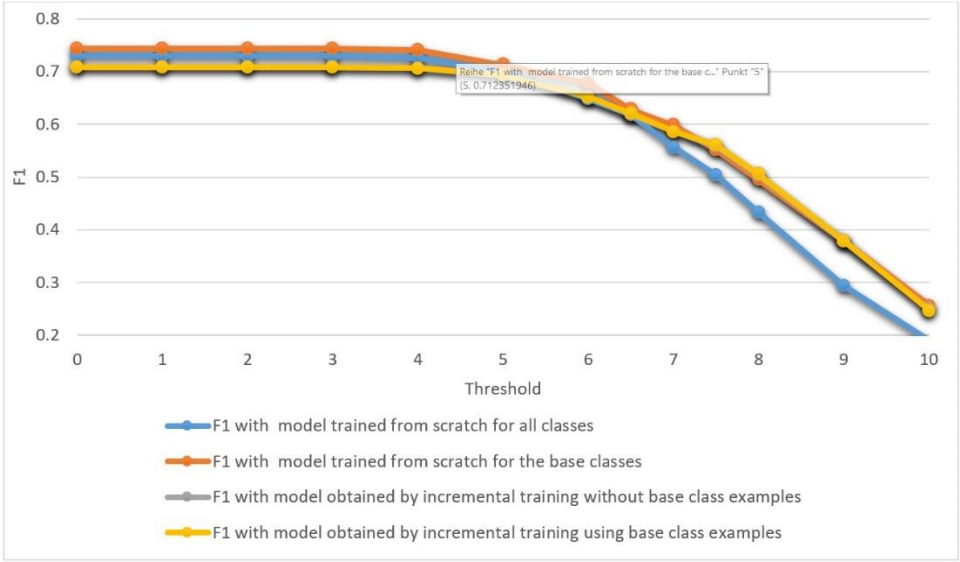


Figure 4-7 Classification result of the 761 base class images (one incremental class)

The diagram shows the classification results for the cases where the model was trained with all classes from scratch (blue line), for training the base classes from scratch (orange line), for incremental training without using

base class examples (grey line covered by the yellow line), and for incremental training using base class examples to train the incremental classes (yellow line).

4.3 2D similarity descriptors (JRS)

4.3.1 Overview

For supporting similarity search in images and video, we have integrated an extraction service for MPEG Compact Descriptors for Video Analysis (CDVA, ISO/IEC 15938-15), as described in this paper. The extractor supports single images as well as the temporal aggregation of descriptors (Bailer, 2017). The descriptors aim to provide a compact and efficient to match representation of a homogenous video segment (i.e., a shot, or subshot if strong visual changes happen within the shot). The descriptor extraction process samples visually sufficiently different keyframes and extracts three descriptor components: (i) a set of interest point based descriptors, (ii) an aggregation of these descriptors using scalable compressed Fisher vectors (SCFV) and (iii) a neural-network based descriptor, obtained from the final feature layer of VGG16, and robust against spatial transformation due to the use of nested invariance pooling (NIP) (Morere, 2017). The descriptor components are binarised or ternarised for efficient matching. We configure the service to not provide the interest point descriptors, which would require pairwise matching. The binarised descriptors provided can be efficiently matches using Hamming distance.

In order to obtain a descriptor for a video segment, the medoid of the keyframe descriptors is determined as representative of the segment. The other descriptors are then coded as differences to this reference descriptor, or discarded, if the difference becomes small. In order to match descriptors of a pair of segments, the reference descriptors are first matched, and if they have at least some similarity, further descriptors can be matched to obtain a score for the pair of video segments.

We have modified our existing implementation of CDVA in order to support Pytorch models trained with current versions of the framework and perform inference using libtorch. For compatibility, a converted >VGG16 model is currently used, but we plan to perform tests with more recent backbone networks.

4.3.2 Tests for cross-modal matching

As NERFs cannot be described with common descriptors for 3D models, and the source images used to create a NERF are already described with 2D descriptors, if they were processed in the XReco platform, we have explored the option to describe a NERF with a set of keyframe descriptors. RAI has provided a set of inference results from different viewpoints from their NERF of Basilica di Superga (see examples in Figure 4-8).

Figure 4-8 Examples of inferences of the Basilica di Superga NERF by RAI.

We have performed a small experiment with data for the Basicila di Superga. We have performed exhaustive pairwise matching (excluding self-matches) between a set of 21 keyframes from RAI videos showing the building, and pairwise matching of keyframes against views rendered from the NERF. The results are shown in Figure 4-9.

Only the efficient neural network-based component of the descriptor has been used. The obtained distributions are quite similar, providing evidence that the descriptor might be a useful tool for cross-modal matching between natural images and NERF views.

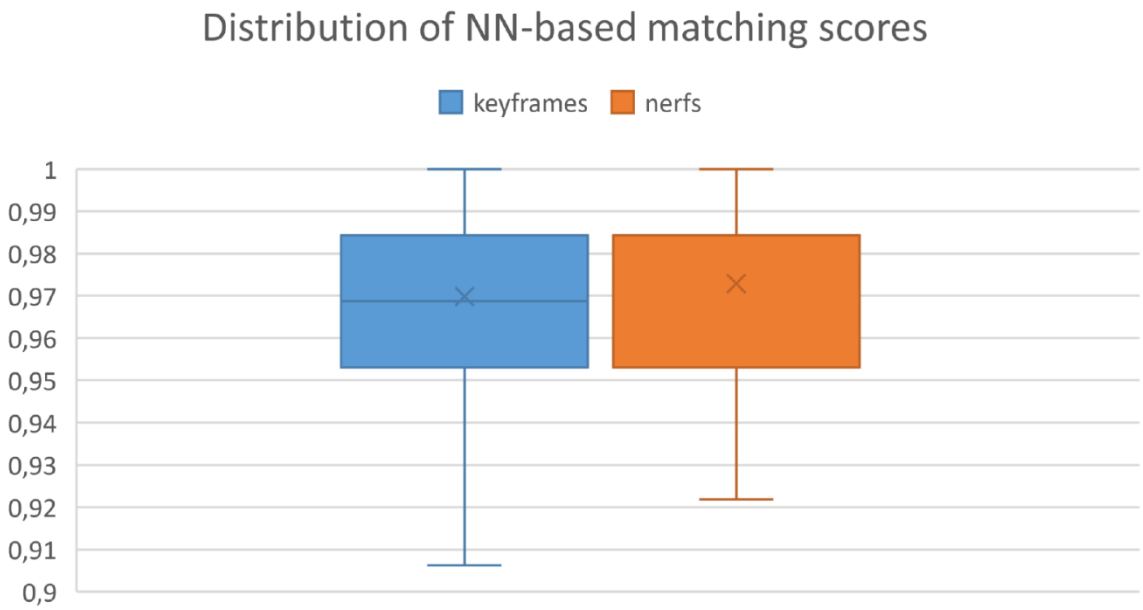


Figure 4-9 Distribution of matching scores between keyframes and rendered NERF views of the same building.

4.3.3 Service

We have implemented a service with a REST interface (see Annex II for the API). The service takes the URL of an image or a video as an input, and in the latter case also the time range to be processed. The service runs the extraction algorithm and returns one or more descriptors for the image/video segment. The extraction process for the neural network-based descriptor components benefits from running on a GPU. The descriptors are 512 element vectors, binarised and designed to be matched using Hamming distance.

4.4 Cross-modal descriptors (CERTH)

The cross-modal retrieval framework aims to provide a comprehensive solution for cross-modal queries. In essence, it facilitates the process of querying information from one modality and obtaining results from a different modality. This innovative approach retrieves data in a seamless and flexible manner, transcending traditional boundaries and opening a world of possibilities for diverse applications (e.g., video captioning, 3D models retrieval). Whether one is seeking to retrieve 3D data results from text queries or images from text inputs, the cross-modal retrieval framework provides a versatile solution for multimodal data access. Specifically, this task supports modalities including, image, mesh, point cloud, and text.

In the following subsections, we present state-of-the-art methods for capturing 3D and multimodal retrieval (Section 4.4.1) and explain our choice of MuseHash (Pegia2023), a multimodal hashing retrieval method, and how we have adapted 3D representations to it (Section 4.4.2). In this deliverable, we showcase results obtained from image, mesh and point cloud modalities (Section 4.4.3). This is primarily because publicly available datasets

include only these modalities. As a future step, we will explore alternatives to showcase datasets that align with all the desired modalities for the task.

4.4.1 Related Work

In this section, we provide an overview of the latest methods in the multimodal object retrieval domain. It covers techniques for image and 3D retrieval. We divide the section into two scenarios: the unimodal and the multimodal scenario.

Unimodal Scenario

Unimodal image retrieval involves using a single modality for retrieval, usually relying on a singular type of data, such as an image. In our research, we investigate different modalities to explore various retrieval scenarios.

Image: Represents the visual content of images, including features extracted from pixels or other image representations. These features capture the visual characteristics of an image, such as shape, color and patterns.

Text: Involves the textual information associated with images, such as captions, keywords, or descriptions, which can be used for retrieval on textual - based queries.

In addition, for specific type of objects there could be different types like these from 3D retrieval and have these modalities (Gezawa, et al. 2020).

Multi-view images: It involves utilizing 2D images for the reconstruction of 3D objects, as outlined in (Su et al., 2015) and (Su et al., 2018). Those methods is beneficial for obtaining diverse viewpoints; however, the quality of the images depends on the captured views. In a recent study (Lin et al., 2022), the authors introduced two self-attention modules, the View Attention Module and the Instance Attention Module, to create a representation of a 3D object by combining three types of features: the original, the view-attentive, and the instance-attentive features.

3D meshes: Surface geometry is defined by vertices, edges, and faces, as explained in (Su et al., 2015) and (Su et al., 2018). While these representations are used in graphics and design, they bring computational and storage challenges. In the case of mesh data, the Mesh Neural Network (MeshNet) (Feng et al., 2019) transforms it into a list of faces, calculating two types of information for each face: a spatial vector based on the centre data and a structural vector using centre, normal, and neighbour information. These features are then combined using a multi-layer perceptron. On the other hand, the Mesh Convolutional Neural Network (MeshCNN) (Hanocka et al., 2019) applies convolution and pooling operations to mesh edges and connected triangles' edges. During pooling, it simplifies edges while preserving the overall mesh structure.

Point Clouds: Describing objects using individual points, especially in robotics (Su et al., 2015; Su et al., 2018), comes with challenges when dealing with sparse and irregular point clouds. To address these challenges, Qi et al. (Qi et al., 2017) introduced Point cloud Network (PointNet), a network architecture designed to effectively handle unordered point clouds and provide a comprehensive end-to-end solution for classification and retrieval tasks. Another approach, the DGCNN proposed by (Wang et al., 2019), uses dynamic graph convolution for processing point cloud data, but it still faces difficulties due to the sparse and irregular nature of the points.

Each representation method presented has its advantages and disadvantages. Multi-view images use multiple perspectives for reconstruction, but accuracy relies on the captured views. Point clouds are storage efficient and precise but come with challenges related to sparsity and irregularity. 3D meshes capture intricate shapes and

details, yet they demand computational resources and storage. Choosing among these representation methods depends on factors such as accuracy, efficiency, and suitability.

In our research, we focus on mesh data for the unimodal scenario as it excels in providing rich information representation, as highlighted in the study by Jing et al. (2021). For our unimodal experiments, we have selected MeshNet and MeshCNN as our preferred choices, guided by the research findings of Jing et al. (2021).

Multimodal Scenario

In the field of multimodal image retrieval, researchers have explored various strategies that involve combining different modalities like visual and textual. Hashing methods have gained attention due to their speed and efficiency in memory usage. We have specifically chosen supervised hashing categories, which include methods based on similarity, adversarial approaches, deep neural networks, and discrete methods.

Fast Cross-Modal Hashing (FCMH) (Wang et al., 2021) uses an auxiliary variable to approximate binary codes, making it possible to optimize these codes by minimizing quantization errors. Discrete Online Cross-modal Hashing (DOCH) (Zhan et al., 2022) is a technique for generating high-quality hash codes across various data types. This is achieved by harnessing both data point similarities and detailed semantic information. Moreover, Label-Attended Hashing (LAH) (Xie2020) takes a unique approach that involves the separate generation of image representations and the labelling of co-occurrence embeddings, incorporating hash functions based on the Cauchy distribution. This allows LAH to effectively capture image-label relationships, thus improving the retrieval performance. Hadamard Codebook based Online Hashing (HCOH) (Lin et al., 2018) utilizes a Hadamard matrix to minimize the l_2 difference between hash-like output and target hash codes, referred to as Hadamard loss. It simultaneously trains both the classification loss and Hadamard loss.

Additionally, there are supervised methods that employ adversarial learning and knowledge transfer. The Self-Supervised Adversarial Hashing Network (SSAH) (Li et al., 2018) combines a self-supervised semantic network with multi-label information. It employs adversarial learning to enhance semantic relevance and ensure feature distribution consistency across different modalities. On the other hand, the Generalized Semantic Preserving Hashing (GSPH) (Manda et al., 2018) approach simultaneously learns optimal hash codes for two modalities and hash functions to map features to those hash codes. Regarding the Matrix Tri-Factorization Hashing Framework (MTHF) (Liu et al., 2019), it focuses on transferring knowledge from a single-modal source domain to a cross-modal target domain, thereby improving cross-modal retrieval. Moreover, Jiang et al. (2018) proposed the kernel-based discrete latent factor model cross-modal hashing (KDLFH) (Jiang 18), which is a discrete method capable of directly learning binary hash codes without continuous relaxation, using a stochastic learning strategy. Lastly, the Central Similarity Quantization (CSQ) (Yuan et al., 2020) introduces a global central similarity concept, encouraging hashing codes for similar images to converge towards corresponding centres.

While many of the mentioned methods are primarily designed for cross-modal scenarios, where queries are typically unimodal even in architectures with multiple modalities, exceptions like Label-Attended Hashing (LAH) support multimodal queries. In the 3D retrieval domain, there's a recent approach called Cross-Modal Center Loss (CMCL) by Jing et al. (2021) that aims to unify point clouds, meshes, and multi-view images in a single framework for multimodal retrieval. In this framework, various 3D modalities are collectively trained to develop representations and identify optimal features, with the aid of loss functions like cross-entropy and mean-square-error to enhance performance. However, it is important to note that CMCL can be computationally intensive due to the integration of multiple 3D modalities into a single framework, potentially requiring substantial computational resources. Additionally, CMCL's performance may vary depending on the dataset, as it is sensitive to the central characteristics of each modality.

4.4.2 Methodology

In this section, we study and present a method designed for retrieval that can be adapted to different scenarios, including unimodal (using only one modality), cross-modal, and multimodal scenarios. Before delving into the specifics of the method, we first define the necessary variables that will be employed in our analysis.

To formally address this problem, we establish the following scenario: We have a query object, denoted as Q, and a database, denoted as DB, containing a collection of 3D objects represented using various views, such as images and meshes and the primary objective is to perform efficient retrieval. This retrieval process aims to find objects within DB that exhibit similarities with the query Q. This involves a thorough examination of the unique features that define Q, followed by a comparison of these features with the corresponding attributes of objects in DB to identify relevant matches.

In the context of 3D retrieval methods, MeshNet and MeshCNN are both unimodal, focusing on mesh-based techniques. However, CMCL takes a different approach as a cross-modal 3D retrieval method. Specifically, it simultaneously learns a shared space for various features from different sources, using MeshNet, DGCNN, and ResNet for mesh, point-cloud, and image modalities, respectively.

In our adaptation of image retrieval techniques for 3D retrieval, we incorporate supervised hashing methods into the architecture to generate hash code features. Specifically, we have opted for LAH and MuseHash, which are known for their effectiveness across various data types, as highlighted in recent research (Pegia et al., 2023). LAH, originally designed for unimodal image retrieval, acquires hash codes by applying a non-linear hash function to mesh features using data from MeshNet. MuseHash (Pegia et al., 2024), on the other hand, utilizes the same models employed by CMCL to extract features from all modalities. Subsequently, MuseHash employs Bayesian ridge regression to learn hash functions, mapping feature vectors to the Hamming space. This allows for the support of both unimodal and multimodal queries.

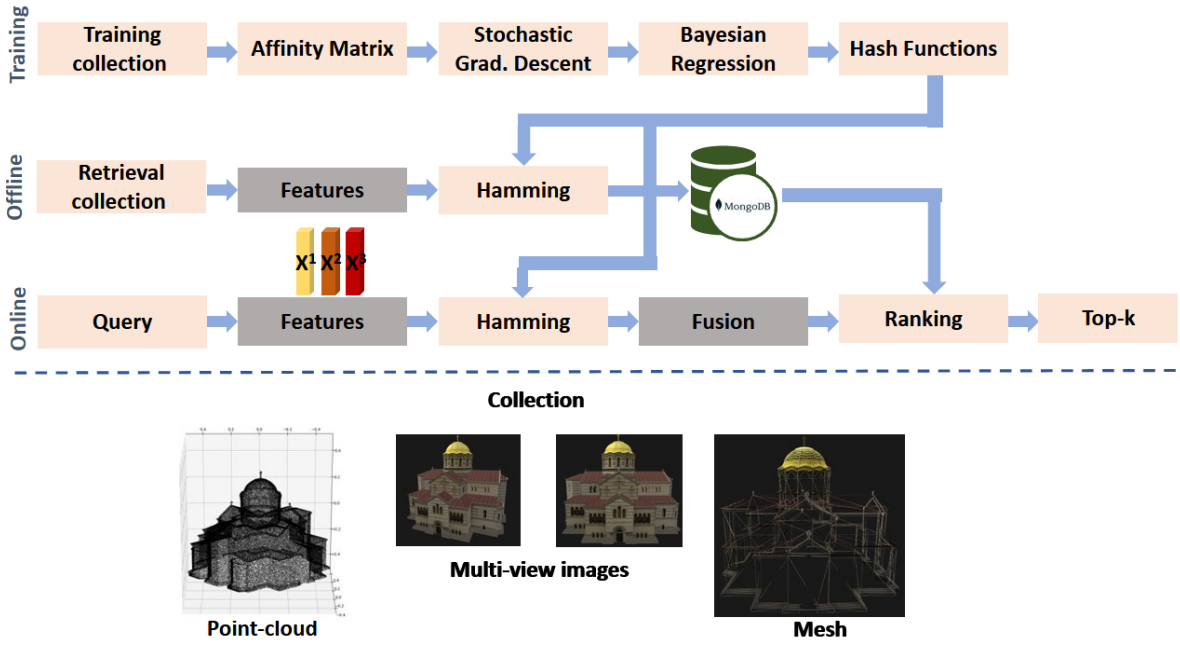


Figure 4-10: Proposed MuseHash framework.

The MuseHash method was adapted in XReco, significantly improving the system's performance in cross-modal and multimodal retrieval tasks. The modified components in the MuseHash architecture are depicted in gray

color. The MuseHash framework, as shown in Figure 4-10, involves two main phases: an offline indexing phase and an online querying phase. During the offline phase, the framework learns hash functions for each modality using Bayesian ridge regression and creates hash codes for all images in the retrieval set. In the online phase, the system extracts features from each modality of the query image, calculates the corresponding hash codes, combines them into a single feature vector, and queries the hash code database. The results are ranked, and the top k relevant results are provided. Importantly, the framework optimizes efficiency by sharing feature extraction structures and learned hash functions between both phases.

In the following, we describe each phase separately to understand the principles governing the MuseHash framework. In order to dive into the details of MuseHash, we give the following notations.

We have an affinity matrix A of size $n \times n$ and learned hash codes H, which are binary codes of length d_c for the training set. Each instance in H represents the projection of a corresponding training set instance I_i. We use U_M to denote the set of learned hash functions for M modalities and u_m^k to represent the learned hash function for the m-th modality and the k-th bit, where k ranges from 1 to d_c. For each modality, c_m is the hash code, and c_k^m is the k-th bit of that code, where k also ranges from 1 to d_c. The hash codes for all M modalities, denoted as $(c^1, c^2, \dots, c^M)$, are combined into a single feature vector, as described below. In MuseHash, the hash code $H_{q,m}$ for a given query q and modality m can be computed as $H_{q,m} = sign(u \times X^m)$, where the operator "$sign$" represents the inner product between vectors.

In the training phase, MuseHash starts by calculating the affinity probabilities from the training labels and converts them into semantic probabilities. Then, it generates feature vectors for each modality and employs Bayesian regression, incorporating the semantic probabilities, to learn the hash functions. These hash functions aid in projecting the features into the Hamming space. Below, one can find the details for each step.

MuseHash starts by using the training labels to create an affinity matrix A, where each element

$$A_{i,j} = \frac{\langle L_i, L_j \rangle}{\|L_i\| \, \|L_j\|} \quad (1)$$

is calculated as the dot-product of two vectors, L_i and L_j, normalized by their magnitudes. Then, it transforms these affinity probabilities into the semantic space P by averaging them, resulting in

$$p_{i,j} = \frac{A_{i,j}}{\sum_{k=1}^{n} \sum_{m=1}^{n} A_{k,m}} \quad (2)$$

These semantic probabilities are mapped into the Hamming space using Kullback-Leibler (Erven et al., 2014) divergence and stochastic gradient descent (Equation (3)). Then, MuseHash applies Bayesian regression to learn a hash function for each modality. Each feature vector is associated with its corresponding hash function.

$$\Psi = \min_{\hat{H} \in \mathbb{R}^{n \times d_c}} \sum_{s=1}^{n} \sum_{t=1, s \neq t}^{n} p_{s,t} \log\left(\frac{p_{s,t}}{q_{s,t}}\right) + \frac{a}{C} \left\| \left| \hat{H} \right| - I \right\|_2^2$$

$$q_{s,t} = \frac{\left(1 + \left\|\widehat{H}_{s,.} - \widehat{H}_{t,.}\right\|_2^2\right)^{-1}}{\sum_{k=1}^{n}\sum_{m=1,m\neq k}^{n}\left(1 + \left\|\widehat{H}_{k,.} - \widehat{H}_{m,.}\right\|_2^2\right)^{-1}}$$

(3)

During the offline phase, MuseHash captures features from the retrieval set for each modality and generates hash codes using the learned hash functions. Each item in the retrieval set is linked to hash codes for each available modality, making it easier to store and retrieve multimedia data.

Specifically, MuseHash collects features $(X^1, X^2, \dots, X^M)$ from the retrieval set, and computes the hash codes $(c^1, c^2, \dots, c^M)$ using the previous learned hash functions. These hash codes are then stored in a database (e.g., MongoDB) for efficient multimedia data storage and retrieval.

Finally, in the online phase, MuseHash uses the hash functions learned earlier to create hash codes for a given query. A query can include one modality or a combination of modalities. These query-specific hash codes are then combined with the stored hash codes, helping the system efficiently locate and retrieve multimedia data that matches the query.

In particular, when dealing with a query instance Q, MuseHash extracts features from each modality, calculates hash codes using the previously learned hash functions, and combines them into a single unified hash code using Equation (4).

$$C_q = f\left(H_{q,1}, H_{q,2}, \dots, H_{q,M}\right) = \sum_{s=1}^{n}\sum_{t=1}^{n} H_{q,t} \oplus H_{q,s}$$

(4)

The symbol $\oplus$ denotes the XOR operator between hash codes through all binary codes.

This unified hash code is then employed to query the database for the top-k relevant results using the Euclidean distance.

4.4.3 Experiments

This section provides information about the datasets used for comparing state-of-the-art methods. It also introduces the experimental settings and presents the results for various retrieval scenarios, including unimodal, cross-modal, and multimodal retrieval. Additionally, the section includes an analysis of the training process.

Datasets

The evaluation of our method and the comparison with existing state-of-the-art methods is done on the two publicly available datasets. We selected these datasets because of their relevance to Demonstrator 1 (News Media) and Demonstrator 2 (Tourism & Automotive).

BuildingNet_v0 The BuildingNet_v0 (Selvaraju, 2021): It offers comprehensive annotations covering a wide range of building types, including structures like churches and palaces, and maintains high-quality standards.

ModelNet40 The ModelNet40 (Wu, 2015): It is an extensive dataset of 3D CAD models, encompassing a diverse array of object categories, such as cars, bottles, and more.

Table 13 contains information on the selected datasets including their collection size, training set dimensions, testing set dimensions, retrieval scope, and the number of distinct labels associated with each dataset.

Table 13: Two benchmark datasets used in experiments

Dataset	Ground Truth Labels	Collections Sizes			
		Whole	Retrieval	Training	Testing
BuildingNet_v0	60	2000	1900	500	100
ModelNet40	40	12311	11696	4843	615

Figure 4-11 depicts examples of images from the two datasets. Specifically, the first row includes an element from the ModelNet40, while the second row an example from the Buildingnet_v0 dataset, respectively. It should be noted that for each element, the mesh, point cloud and image modality are presented.

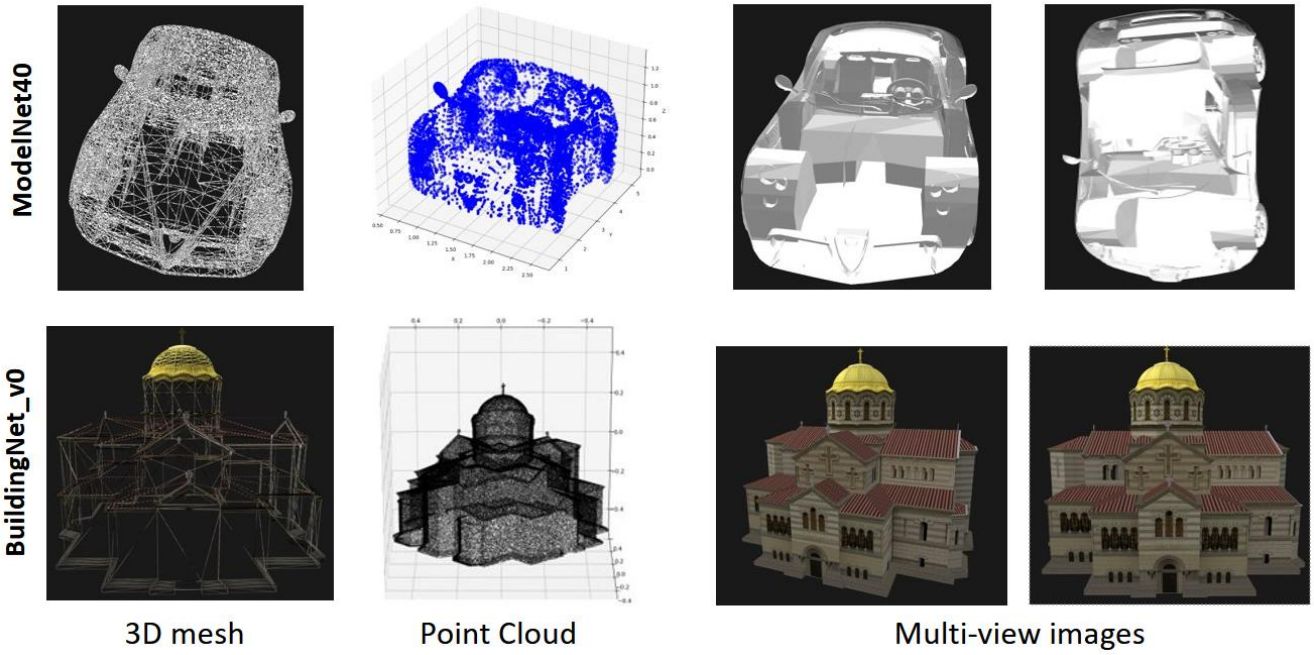


Figure 4-11: Some images from ModelNet40 and BuildingNet_v0 datasets.

Experimental Settings

In our experiments, we assess the performance of two distinct sets of methods using a variety of metrics. Specifically, we focus on evaluating the hashing methods, MuseHash (Pegia et al., 2023) and LAH (Xie et al., 2020), while examining the impact of different hash code lengths ($d_c = 16, 32, 64, 128$). For each volumetric method, we experiment with the number of epochs ($epoch = 10, 50, 100, 150$) used for metric computation. We follow the recommended training and testing sizes as suggested by the dataset authors (Wu, 2015; Selvaraju, 2021).

In our experiments, we converted various modalities into feature vectors to make them compatible with the hashing methods. For the visual modality, we conducted an averaging process that involved 180 multi-view image feature vectors extracted from the fc-7 layer of ResNet50, resulting in a 2048-D vector. In the case of point

cloud and mesh modalities, we directly obtained 256-D vectors from the final layers of DGCNN and MeshNet, respectively.

We compare our approach with two leading 3D mesh methods, the MeshNet (Gezawa, 2020) and the MeshCNN (Hanocka2019), as well as a cross-modal 3D retrieval method, CMCL (Jing, 2021). Our comparison is based on various metrics, including mean Average Precision (mAP), precision at k (prec@k), recall at k (recall@k), f-score at k (fscore@k), accuracy, and training time. The metrics are defined in Annex IV.

In all experiments, we used a 5-fold cross-validation methodology to ensure a more robust evaluation, and the runtime of experiments was measured per epoch or per hash code length.

Additionally, it is important to note that all 3D retrieval implementations consumed significant amounts of memory, while MuseHash operates with minimal memory usage, utilizing only a small number of bits.

Unimodal Retrieval Results

In our study, we place a significant emphasis on utilizing mesh data in a unimodal context because of its superior performance, owing to its ability to provide rich information representation (Jing et al., 2021). Specifically, for our unimodal experiments, we have selected the latest techniques, MeshNet and MeshCNN, as our preferred options based on the research findings (Jing et al., 2021). The results of those methods over the two datasets can be found in Figure 5 and Figure 6 for mAP and accuracy, respectively. These charts illustrate the performance of 3D retrieval methods, including MeshNet, MeshCNN, and CMCL, across various values of training epochs..

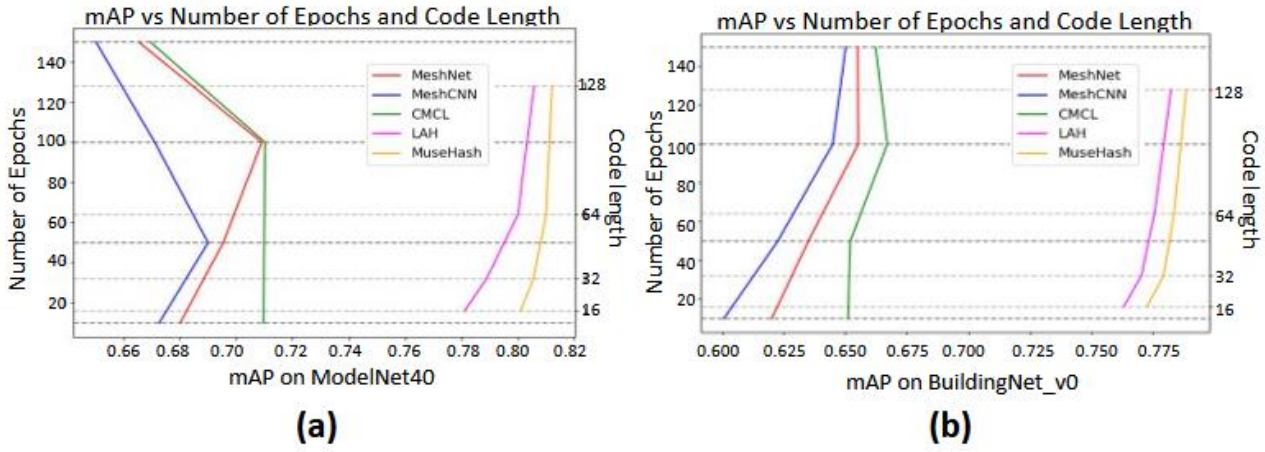


Figure 4-12: MAP results on ModelNet40 and BuildingNet_v0 with different code lengths or number of epochs and mesh modality.

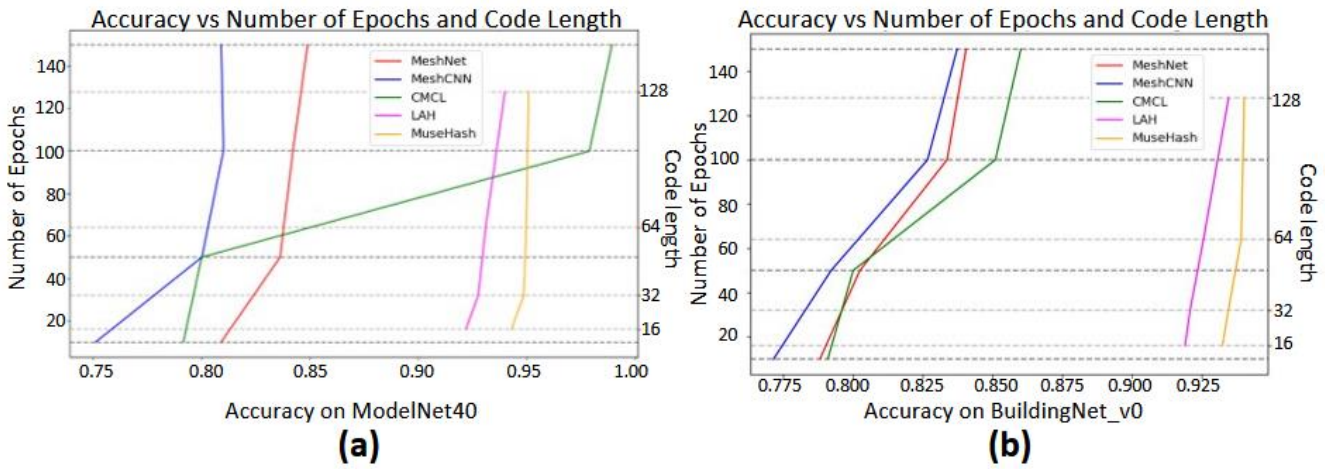


Figure 4-13: Accuracy results on ModelNet40 and BuildingNet_v0 with different code lengths or number of epochs and mesh modality.

The graph's x-axis represents the mAP values, serving as an indicator of the performance of these methods. On the left-hand side of the chart, you can observe the number of training epochs, offering insights into the models' training duration and stability. Simultaneously, the right side of the chart displays the various code lengths used, reflecting the granularity of feature representations. The dashed lines in black and grey correspond to the values of each method at a specific epoch or code length, respectively. Notably, the MuseHash algorithm outperforms the other methods across both datasets, demonstrating superior performance across different hash lengths and epochs.

Furthermore, CMCL approach attains the highest accuracy on the ModelNet40 dataset with an increased number of epochs. However, its mAP performance falls short in comparison. This suggests that CMCL excels in classification tasks but may encounter difficulties when it comes to organizing retrieval results effectively. Moreover, it is important to highlight that some image retrieval methods exhibit better performance in the context of 3D retrieval tasks when compared to the current state-of-the-art 3D retrieval methods.

Cross-modal Retrieval Results

In this section, we focus on cross-modal results where one type of data is used as a query to retrieve information from a different type of data. We explore all six possible combinations of these scenarios. the methods under evaluation, CMCL, LAH, and MuseHash exhibit the capability to perform these cross-modal queries, making it possible to search for information across various data modalities.

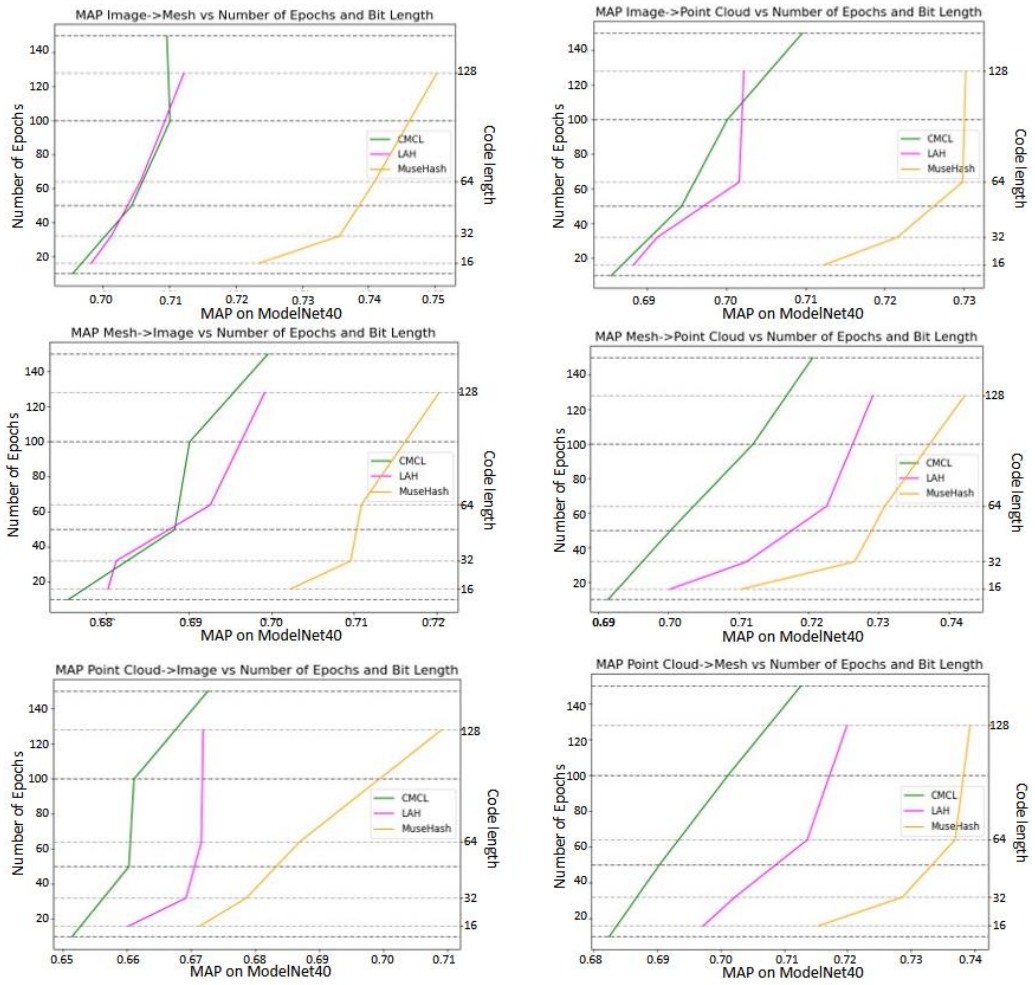


Figure 4-14: Cross-modal results in terms of MAP on ModelNet40 dataset.

Figure 4-14 and Figure 4-15 show the mAP values for each dataset. Among these methods, MuseHash stands out as the top performer, particularly on the ModelNet40 and BuildingNet_v0 datasets. More precisely, MuseHash achieves the highest mAP values on the ModelNet40 dataset (Figure 4-14), with a score of approximately 0.75, in both "Image to Mesh" and "Mesh to Point Cloud" cases. Regarding the BuildingNet_v0 dataset (Figure 4-15), a similar pattern of performance is observed, with slightly lower mAP values, around 0.74. This indicates that MuseHash excels at finding the right information, especially when searching across diverse data types like images and 3D models. Moreover, regarding the BuildingNet_v0 dataset MuseHash produces comparatively lower results, possibly due to the complexities in inferring material information within the 3D data.

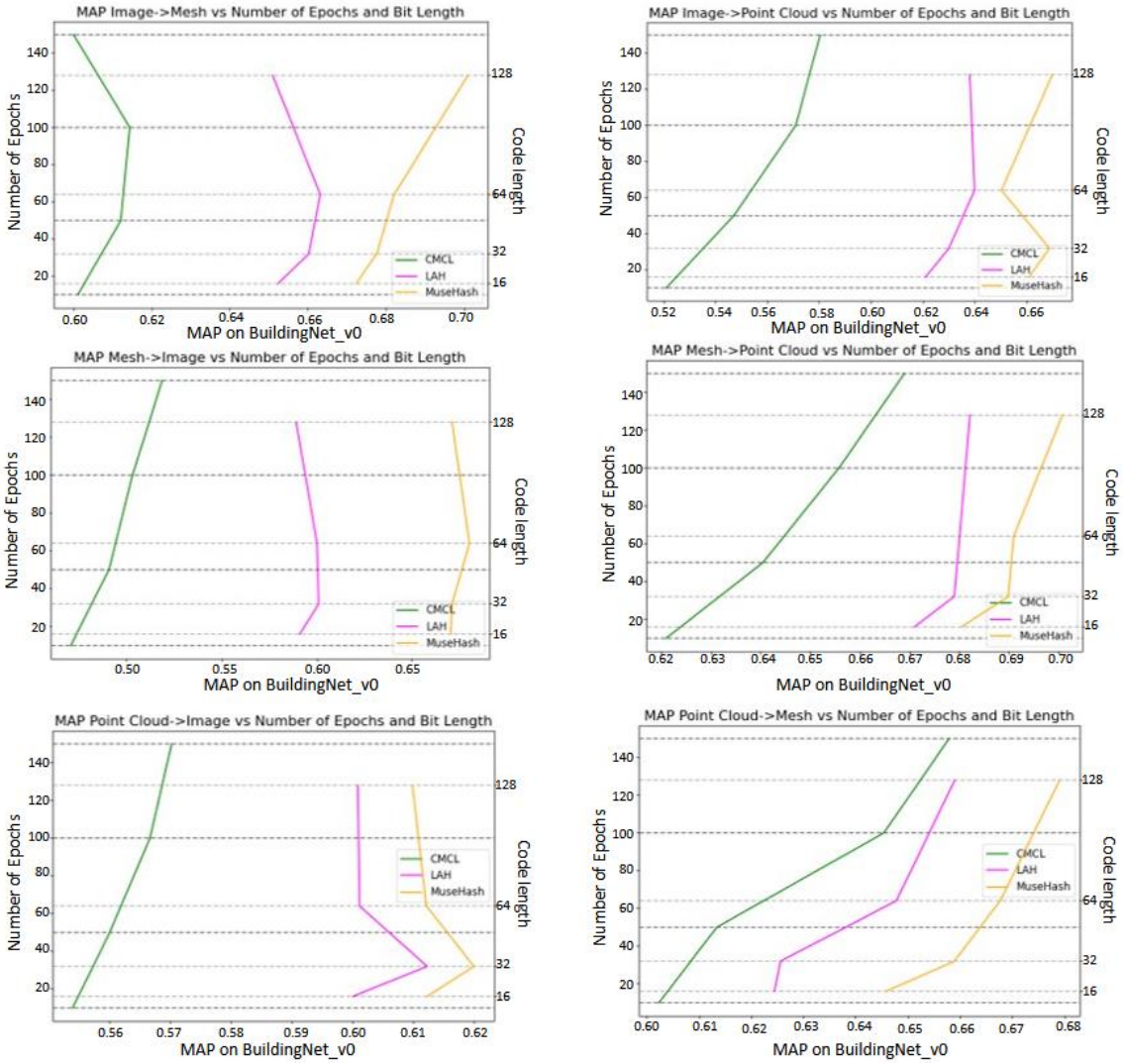


Figure 4-15: Cross-modal results in terms of MAP on BuildingNet_v0 dataset.

Multimodal Retrieval Results

In the context of multimodal scenarios, we investigate the simultaneous use of point clouds, meshes, and multi-view images. The outcomes of these approaches, concerning both mAP and accuracy, are presented in Figure 4-16 and Figure 4-17 for the ModelNet40 and BuildingNet_v0 datasets. Figure 4-16 provides information on the mAP values for MuseHash and LAH for varying code lengths. It also illustrates how the performance of CMCL varies with the number of training epochs. To clarify, the x-axis indicates the method's performance, the left side represents the training epochs, and the right side indicates the code lengths for feature representation.

The results indicate that MuseHash tends to exhibit better performance as the code lengths become longer, reaching its peak performance under these conditions. MuseHash achieves slightly improved mAP values when dealing with multimodal scenarios (Figure 4-12) in both datasets compared to unimodal (Figure 4-16) and cross-modal (Figure 4-15) scenarios. In the case of the ModelNet40 dataset, all methods show improved results with fewer training epochs or shorter code lengths, which is in contrast to the BuildingNet_v0 dataset.

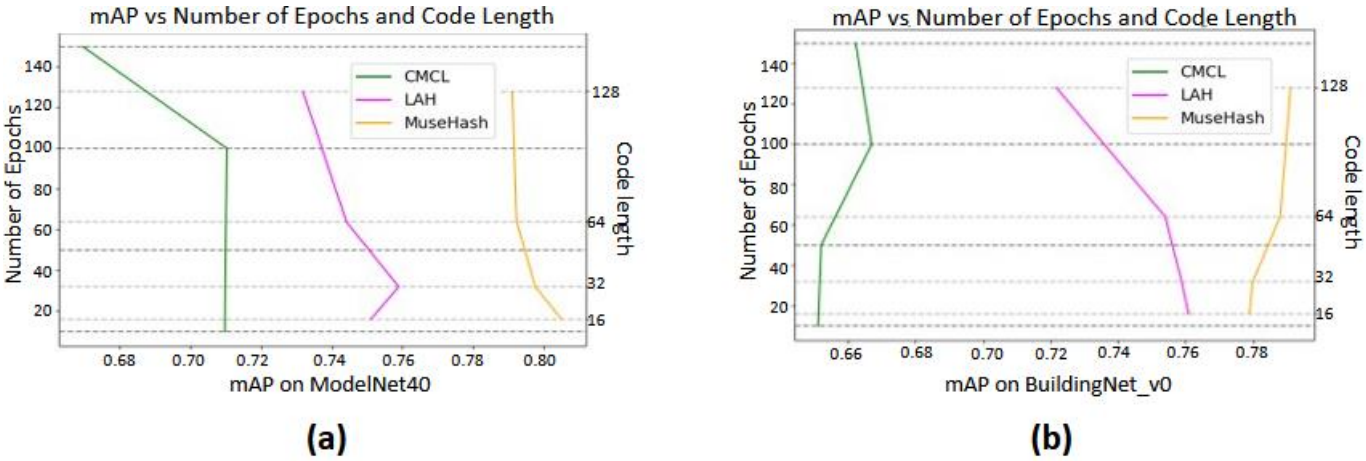


Figure 4-16: MAP values on multimodal scenario.

Similarly, in the Unimodal Scenario, CMCL reaches the highest value in terms of accuracy followed by MuseHash. However, the mAP value of MuseHash is higher than CMCL.

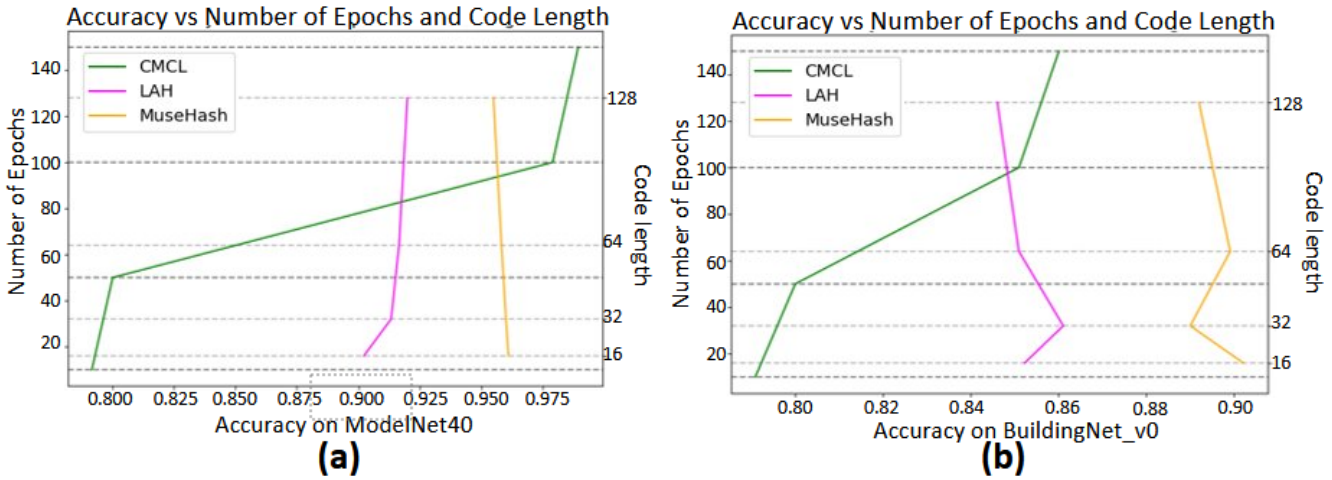


Figure 4-17: Accuracy values on multimodal scenario.

Table 14 and Table 15 give a detailed look at different methods using Precision@k, Recall@k, and Fscore@k with k values of 10, 25, and 50. The metrics are computed for various various code lengths or epochs on the ModelNet40 and BuildingNet_v0 dataset, showing how well these methods rank and fetch relevant items. The chosen metrics help us understand how these methods rank items and their ability to grab important data in the top-k results. As far as the methods being evaluated they can be grouped into two sets. The first involves the methods MeshNet, MeshCNN, and CMCL that are assessed across various epochs, while the second includes the three predefined versions of the MuseHash method (MuseHash1, MuseHash2, MuseHash3) and the LAH method, tested with different code lengths.

Table 14: Comparison of all methods based on Precision at k (k = 10, 25, 50) for different number of epochs or code lengths on ModelNet40 dataset.

Method		Variable Epochs	Precision@k			Recall@k			Fscore@k		
			10	25	50	10	25	50	10	25	50
MeshNet		10	0.6510	0.6560	0.6410	0.6802	0.6533	0.6602	0.6653	0.6546	0.6500
		50	0.6810	0.6712	0.6678	0.7011	0.7051	0.7187	0.6909	0.6877	0.6923
		100	0.6901	0.6854	0.6802	0.7029	0.7011	0.7089	0.6964	0.6932	0.6943
		150	0.7010	0.6910	0.6824	0.7091	0.7123	0.7189	0.7055	0.7015	0.7002
MeshCNN		10	0.5822	0.5701	0.5623	0.5791	0.5607	0.5689	0.5806	0.6011	0.6178
		50	0.6001	0.5803	0.5734	0.5998	0.6011	0.6183	0.5999	0.5905	0.5950
		100	0.6245	0.6183	0.6002	0.6011	0.6190	0.6189	0.6126	0.6186	0.6094
		150	0.6221	0.6112	0.6009	0.6005	0.6123	0.6230	0.6111	0.6117	0.6118
CMCL		10	0.8290	0.7679	0.7142	**0.9985**	0.9943	0.9968	0.9011	0.8666	0.8321
		50	0.8291	0.7687	0.7147	0.9883	0.9943	**0.9968**	0.9018	0.8671	0.8325
		100	**0.8298**	0.7687	0.7149	0.9894	**0.9944**	**0.9968**	**0.9019**	0.8671	0.8326
		150	0.8283	0.7677	0.7142	0.9865	**0.9944**	**0.9968**	0.9013	0.8665	0.8322
Method		Code Length	10	25	50	10	25	50	10	25	50
LAH		16	0.6190	0.6179	0.6242	0.9215	0.9243	0.9268	0.7405	0.7407	0.7460
		32	0.6202	0.6287	0.6347	0.9383	0.9343	0.9461	0.7468	0.7516	0.7597
		64	0.6298	0.6287	0.6349	0.9584	0.9444	0.9468	0.7601	0.7549	0.7601
		128	0.6281	0.6271	0.6242	0.9265	0.9344	0.9468	0.7487	0.7505	0.7524
MuseHash	1	16	0.6412	0.6501	0.6623	0.9567	0.9689	0.9781	0.7454	0.7781	0.7898
		32	0.6589	0.6620	0.6778	0.9612	0.9723	0.9612	0.7818	0.7877	0.8018
		64	0.6601	0.6789	0.6801	0.9967	0.9712	0.9789	0.7845	0.7992	0.8026
		128	0.6791	0.7123	0.7256	0.9701	0.9734	0.9601	0.7989	0.8229	0.8265
	2	16	0.6671	0.6810	0.7020	0.9612	0.9723	0.9865	0.7806	0.8010	0.8203
		32	0.6910	0.7001	0.712	0.9546	0.9612	0.9667	0.8017	0.8101	0.8195
		64	0.7662	0.7405	0.7156	0.9712	0.9781	0.9801	0.8566	0.8429	0.8272
		128	0.8010	**0.8588**	**0.8423**	0.9865	0.9902	0.9923	0.8841	**0.9198**	**0.9112**
	3	16	0.6480	0.6501	0.6589	0.9523	0.9678	0.9621	0.7712	0.7778	0.7821
		32	0.6510	0.6678	0.6781	0.9678	0.9698	0.9512	0.7784	0.7910	0.7918
		64	0.6782	0.6789	0.6834	0.9701	0.9700	0.9634	0.7983	0.7988	0.7996
		128	0.7012	0.6910	0.6901	0.9701	0.9623	0.9603	0.8140	0.8044	0.8038

After careful observation of the tables, it is clear that in general multimodal approaches perform better than MeshCNN and MeshNet in both datasets. This points out the limitations in the architecture or feature representation of the latter two when exclusively working with the mesh view. While CMCL may occasionally outperform, MuseHash is a more efficient option, considering the balance between the performance and the training time. Moreover, MuseHash's ability to combine different modalities, like mesh and image, into a unified hash code improves retrieval accuracy and diversity. Overall, MuseHash is especially valuable in situations with large datasets and resource constraints, where quick and accurate similarity searches are essential.

Table 15: Comparison of all methods based on Precision at k (k = 10, 25, 50) for different number of epochs or code lengths on BuildingNet_v0 dataset.

Method		Variable Epochs	Precision@k			Recall@k			Fscore@k		
			10	25	50	10	25	50	10	25	50
MeshNet		10	0.6012	0.6069	0.6130	0.6520	0.6510	0.6554	0.6256	0.6302	0.6335
		50	0.6120	0.6143	0.6255	0.6670	0.680	0.6701	0.6383	0.6410	0.6470
		100	0.6210	0.6239	0.6301	0.6640	0.6701	0.6723	0.6418	0.6472	0.6505
		150	0.6314	0.6322	0.6367	0.6701	0.6723	0.6731	0.6502	0.6520	0.6544
MeshCNN		10	0.5712	0.5700	0.5620	0.5810	0.5710	0.5701	0.5761	0.5700	0.5660
		50	0.5910	0.5905	0.5723	0.5720	0.5723	0.5811	0.5813	0.5858	0.5767
		100	0.5991	0.5906	0.5813	0.5801	0.5822	0.5759	0.5894	0.5832	0.5786
		150	0.5776	0.5810	0.5792	0.5893	0.5910	0.5882	0.5834	0.5846	0.5837
CMCL		10	0.7120	0.7198	0.7234	0.9013	0.9024	0.9127	0.7955	0.8049	0.8071
		50	0.7121	0.7123	0.7259	0.9122	0.9178	0.9282	0.7998	0.8060	0.8147
		100	0.7239	**0.7265**	0.7201	0.9145	0.9166	0.9201	0.8081	0.8119	0.8079
		150	0.7210	0.7191	0.7145	0.9176	0.9177	0.9145	0.8075	0.8051	0.8022
Method		Code Length	10	25	50	10	25	50	10	25	50
LAH		16	0.6011	0.6045	0.6101	0.9013	0.9050	0.9028	0.7212	0.7241	0.7281
		32	0.6101	0.6123	0.6143	0.9067	0.9104	0.9177	0.7294	0.7345	0.7360
		64	0.6198	0.6205	0.6254	0.9120	0.9134	0.9201	0.7380	0.7412	0.7447
		128	0.6201	0.6243	0.6190	0.9234	0.9100	0.9192	0.7420	0.7436	0.7398
MuseHash	1	16	0.6301	0.6321	0.6350	0.9310	0.9345	0.9407	0.7516	0.7561	0.77582
		32	0.6390	0.6401	0.6429	0.9401	0.9410	0.9451	0.7608	0.7633	0.7652
		64	0.6401	0.6491	0.6510	0.9518	0.9545	0.9561	0.7654	0.7732	0.7746
		128	0.6510	0.6551	0.6670	0.9617	0.9647	0.9670	0.7764	0.7811	0.7895
	2	16	0.6210	0.6280	0.6245	0.9556	0.9560	0.9601	0.7528	0.7593	0.7568
		32	0.6340	0.6359	0.6408	0.9601	0.9623	0.9654	0.7637	0.7667	0.7703
		64	0.6501	0.6523	0.6600	**0.9656**	**0.9669**	**0.9710**	0.7770	0.7804	0.7858
		128	**0.7240**	0.7256	**0.7368**	0.9607	0.9634	0.9656	**0.8257**	**0.8286**	**0.8358**
	3	16	0.6345	0.6340	0.6401	0.9423	0.9456	0.9470	0.7584	0.7595	0.7639
		32	0.6401	0.6422	0.6470	0.9432	0.9423	0.9452	0.7626	0.7648	0.7682
		64	0.6323	0.6301	0.6422	0.9323	0.9333	0.9342	0.7535	0.7526	0.7612
		128	0.6389	0.6401	0.6422	0.9412	0.9410	0.9398	0.7611	0.7615	0.7630

Analysis of Runtime Requirements

Finally, moving to the time performance of the algorithms, Figure 4-18 illustrates the training times, measured in minutes, for several methods, including MeshNet, MeshCNN, CMCL, LAH, and the three different versions of MuseHash, on the ModelNet40 and BuildingNet_v0 datasets. These measurements were recorded across different training epochs and code lengths. In each graph, ach line on the plot corresponds to a specific method, the x-axis indicates the training time in minutes, the left y-axis represents the number of training epochs, and the right y-axis signifies the code length used during the training process. The black and grey dotted lines denote the values of each method for a particular epoch or code length. Specifically, the three MuseHash variants

(MuseHash1, MuseHash2, and MuseHash3) were evaluated with varying code lengths, corresponding to the use of mesh data, mesh data combined with visual information, and mesh data along with visual and point cloud views.

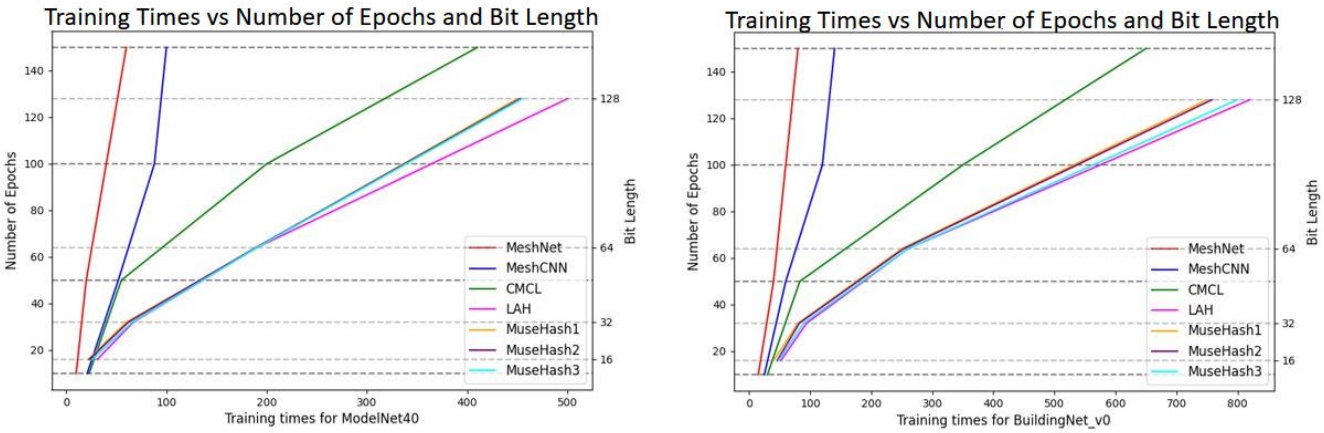


Figure 4-18: Comparison of training times for all methods on ModelNet40 and BuildingNet_v0 datasets in minutes.

Therefore, following thorough observation of Figure 4-18, it becomes evident that in terms of training times, , MeshNet and MeshCNN prove to be relatively faster, making them computationally more efficient choices compared to CMCL and the MuseHash variants. CMCL stands out for requiring significantly more time for the same number of epochs. Among the MuseHash variants, 'MuseHash1' consistently shows the shortest training times across various code lengths, making it the most time-efficient option. However, as the code length increases, all MuseHash variants experience longer training times due to the increased complexity of computations and higher memory requirements. To sum up, a definitive observation drawn from Figure 4-18 is that 'MuseHash1' is the fastest, yet it experiences longer training times as the code length increases.

4.5 2D object detection and tracking (i2CAT)

The information about object classes and their motion may provide useful information for content retrieval in media production. In this section, we will describe the object detection and tracking system of the XReco service, along with the context of the system. The developed system includes two main AI modules: the multi-object detection system and the multi object tracking system. By combining the two modules, we obtain a system capable of performing multiple object detection and tracking (MODT) in 2D. To achieve this, the state of the art in object detection and tracking has been investigated to identify new algorithms and solutions for testing.

With regard to object detection, several state of the art new neural networks have been tested from the literature, YoloV5[11] and YoloX (Ge, 2021) are an example of them, they are iterative evolutions of the existing YOLO (You Only Look Once) based detectors well known in the state of the art for its speed, as it only uses a single convolutional neural network (CNN) to simultaneously generate region proposals and classify these regions. This approach gives to Yolo the capacity to run inferences very fast, achieving real-time speeds.

[11] https://docs.ultralytics.com/yolov5/

- YoloV5: The YoloV5 neural network has a single-stage anchor-based object detector with 3 main components; First, a Cross Stage Partial (CSP) neural network is used to extract rich features from the image. Second, a PANet is used to extract feature pyramids, which helps the model to scale across different object sizes. Finally, a head that applies the anchors to the features to generate the final detections.
- YoloX: Unlike YoloV5 and its Yolo-based predecessors, YoloX is a model not based on anchors, making it simpler than the previous ones but achieving better performance on the COCO dataset. The most representative change in this network is in the head. YoloX implements a decoupled head to separate the classification task (assigning a label or class to an object) of the regression task (predict the position and shape of an object in the image). This approach solves the problems presented by single-head networks, in which the anchors might not be optimal for the problem, resulting in network generalization issues. In addition to this, the use of a decoupled head allows YoloX to slightly reduce inference time compared to other Yolo-based networks.

In Figure 4-19 you can see the difference between YoloV3, Yolov5 and YoloX architectures.

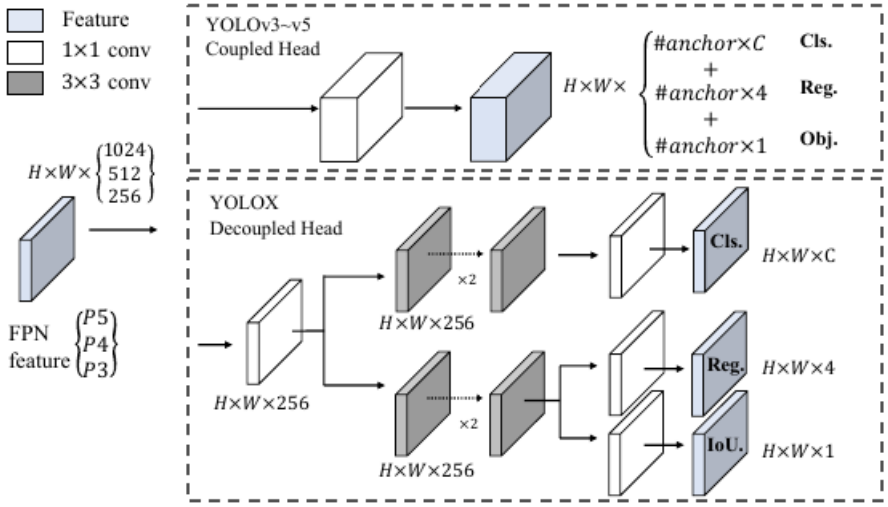


Figure 4-19 Differences between YoloV3, YoloV5 networks and YoloX, which incorporates a decoupled head.

To verify the performance of the algorithms, a series of tests have been carried out on videos recorded on a highway. Figure 4-20 shows some of the results for the ScaledYoloV4 (Wang, 2021), YoloX, and YoloV5 detectors.

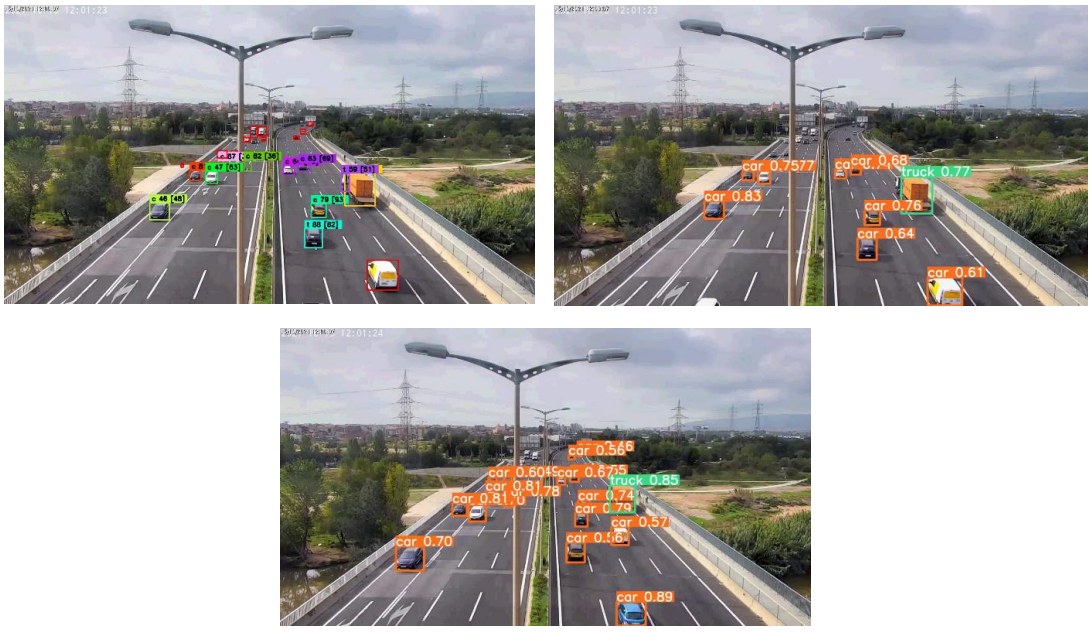


Figure 4-20 Sample detection in one frame with ScaledYoloV4 (top left), YoloX (top right), and YoloV5 (bottom)

As can be seen, the results obtained with the new YoloX and YoloV5 detectors are equivalent to those obtained with ScaledYoloV4, offering greater speed when using smaller image sizes. Even so, it can be seen how in the case of YoloX some of the vehicles in the background, when using a smaller image size, are not always detected as they are too small.

Regarding tracking, research on the state of the art has been conducted where we can highlight the BoT-SORT (Aharon, 2022) tracker. This tracker offers several advantages over other object tracking methods in video sequences. Firstly, the effect of camera shake can be corrected by estimating its position and orientation relative to the scene. Secondly, it improves the accuracy of the Kalman (Kalman, 1960) filter by using a state vector that includes the position, object size (width, height), velocity, and acceleration of the objects. Thirdly, you have the possibility of using object re-identification neural networks to obtain information about the appearance of the objects, which is used together with the motion information obtained from the Kalman filter to have more robust tracking.

Finally, the combination of detection and tracking has also been tested with YoloX (detector) and BoT-SORT (tracker). We have decided to use YoloX because it obtains similar results, and it has Apache-2.0 license[12] instead of proprietary ones such as for Yolo_v5 to Yolo_v8 family. In Figure 4_3_track_res you can see an example result in 2 frames about 1-2 seconds apart in the test video. As can be seen, all the vehicles that have been detected and assigned in the first frame and that are also detected in the second frame have kept the ID, as BoT-SORT has been able to keep the vehicles well identified. Additionally, you can also see how the vehicles that did not appear in the first frame have been assigned a new identifier.

[12] https://github.com/Megvii-BaseDetection/YOLOX/blob/main/LICENSE

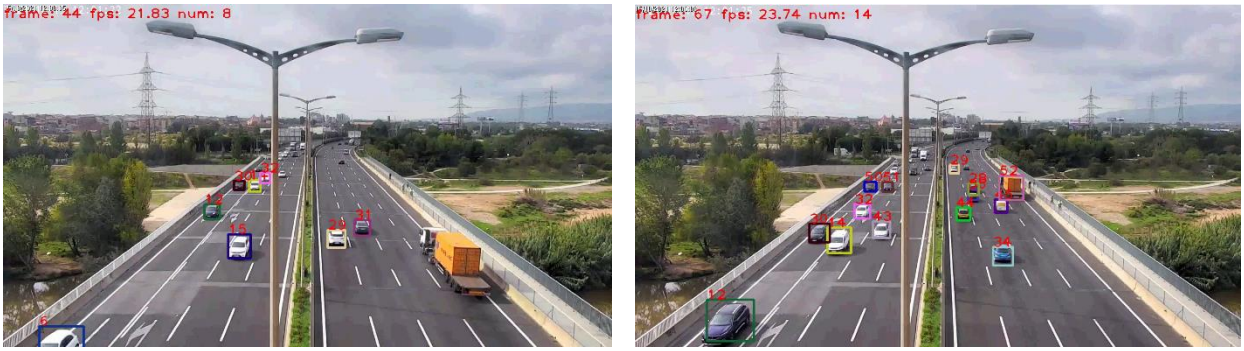


Figure 4-21 Detection and tracking with YoloX and BoT-SORT in 2 frames separated by about 2 seconds

The service's object detection and tracking system incorporates state-of-the-art algorithms, YoloX and Bot-Sort, to achieve accurate and robust object detection and tracking in video sequences.

The 2D Object detection and tracking models can be accessed through a Dockerized API, which provides a secure and isolated environment for running the models and their dependencies. This approach allows for easy deployment and scaling of the application, as well as the flexibility to assign resources based on the specific requirements of each model. A general application diagram can be found in Figure 4-22.

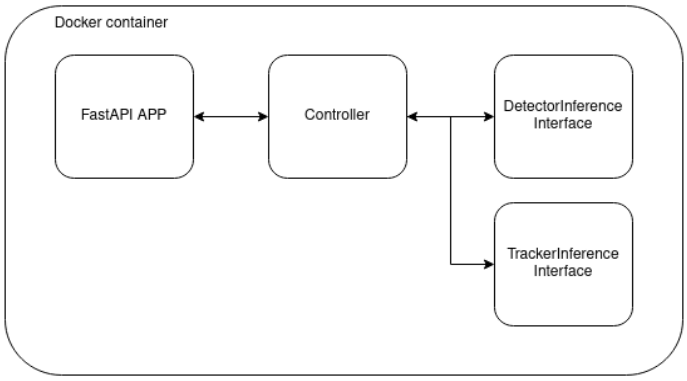


Figure 4-22 Detector and tracker application diagram.

The FastAPI app exposes endpoints to the users and allows making inference request to the application and sending the inference results back to the users. The API endpoints trigger Controller method calls, a class that contains all the application logic and orchestrates all the operations between the AI modules. In order to make the design as algorithm/model agnostic as possible, two interfaces have been implemented.

5 User and workflow management

This section discusses infrastructure functionalities that are needed to implement both the content search functionalities as well as link them to the reconstruction services. One is user authentication, which is particularly relevant if the XReco platform is deployed as a hosted solution with users from multiple parties, but also, when remote services are invoked from the platform. The other are content baskets, as a concept of collecting search results related to a production, select a set of third-party items to be ingested or define a set of items to be passed to a particular reconstruction service.

5.1 Authentication services (Atos)

We faced the challenge of managing secure and efficient user authentication across a distributed microservice architecture. To address this, we deployed Keycloak[13] with JWT[14], a solution that not only enhances security but also aligns with our scalability needs.

JWT is a compact, self-contained way for securely transmitting information between parties as a JSON object. In a microservice architecture, they serve two primary purposes: authentication and information exchange. JWTs are stateless; they contain all the necessary information about the user, eliminating the need for a centralized session storage. This is crucial in a microservice architecture, where services need to scale independently without relying on shared state. The security of JWT stems from its ability to be signed, which ensures that the tokens can be trusted and verified. Additionally, their compact structure makes them suitable for high-traffic networks where a significant amount of requests are made across services.

Keycloak is an open-source identity and access management solution. It provides out-of-the-box support for managing user federation, securing APIs, and enabling Single Sign-On (SSO). One of Keycloak's strengths is its ease of integration with existing applications and services, facilitating a centralized approach to security in a decentralized architecture.

This solution acts as an identity broker and authentication server. It centralizes the login functionality for all services in the architecture. When a user logs in, Keycloak authenticates their credentials against its user store or external sources and issues a JWT. This token is then used to access secured resources across the microservices. The services validate the token using Keycloak's public key, ensuring both the authenticity and integrity of the token.

Keycloak's functionality is largely based on OAuth 2.0[15] and OpenID Connect[16] standards. OAuth 2.0 is the industry-standard protocol for authorization. OpenID Connect is an authentication layer on top of OAuth 2.0. It allows clients to verify the identity of the end-user and to obtain basic profile information in an interoperable and REST-like manner. In the context of Keycloak, OpenID Connect is used for user authentication, complementing OAuth 2.0's focus on authorization.

There are several options for implementing these standards, and we have opted for an easily integrable solution, which is depicted in sequence diagram of the figure (Figure 5-1), and later explained in detail. The diagram represents the Open ID implementation. For services that does not require an user interface, we will enable

[13] https://www.keycloak.org/
[14] https://jwt.io/
[15] https://oauth.net/2/
[16] https://openid.net/

"Resource Owner Password Credentials Grant" for this client, meaning the client service will directly exchange the user credentials with a token.

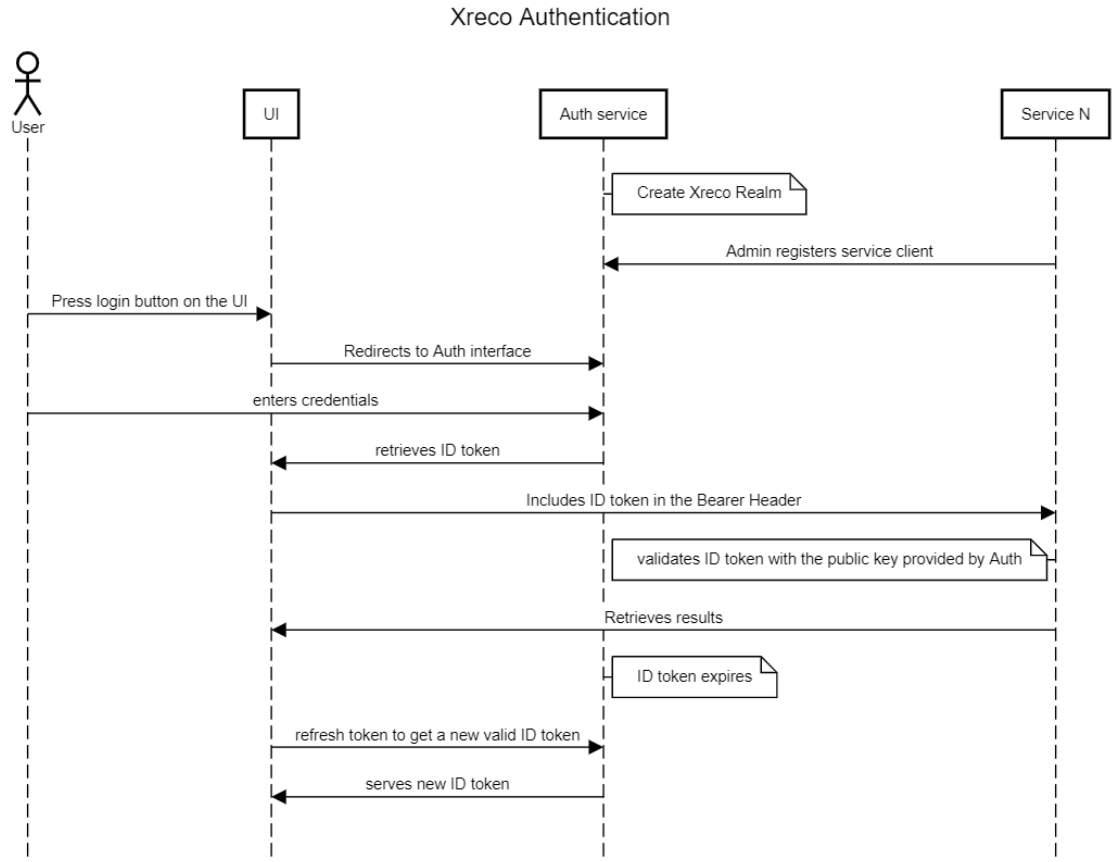


Figure 5-1 Auth implementation sequence diagram

Realm creation: A realm is an abstract space, where the Users, Roles and client configurations are shared. We will create one for XReco

Client Registration: Initially, the service must be registered with Keycloak as a client. This setup includes specifying the client ID, client secret (for confidential clients), allowed redirect URIs, and other relevant settings.

Authentication Request: When a user needs to authenticate, the client application redirects them to Keycloak. This redirection is an OpenID Connect authentication request.

User Authentication: The user logs in directly through Keycloak's authentication interface. For the moment, we simplified this step with username and password, although in the future Keycloak allows several other methods like multi-factor authentication, or even social logins if configured.

Token Issuance: Upon successful authentication, and for the sake of simplicity we implemented the implicit Oauth flow, where the tokens are issued directly to the client.

There are three tokens to consider:

- ID Token: A JWT that contains information about the authenticated user (claims).
- Access Token: Used by the client to access protected resources (the Xreco Microservices) on behalf of the user.

- Refresh Token (optional): Used to obtain new access tokens without requiring the user to re-authenticate.

For the Authorization we simplified the workflow defining different user roles in the platform and including them in the ID JWT token directly. This will allow each service to decode the user token and grant permissions based on the role. The exact accessible services for each role are still to be defined after the first full integration, but we made some testing implementation creating user groups per organization and assigning different roles to these groups or specific users, and the approach to the keycloak framework fits perfectly.

5.2 Content basket management (Atos, UNIBAS)

As a requirement for the project's specific infrastructure, we have implemented a content basket system for users. The transformation services proposed by the project (D4.1) are very hardware-intensive, meaning there are services that will take hours or even days to produce a final result. This is why we created these content baskets where metadata can be stored, including the storage location and the process status, of the content that the user wants to transform.

The infrastructure of the content baskets is based on that of the metasearch. We use Elasticsearch as a non-relational database to store the necessary metadata in JSON format. We also benefit from the implementation of the authorization service in the project, as we extract the user ID directly from the JWT generated for the baskets to assign them to a specific user.

The way to interact with these baskets is defined in the interactive documentation provided by our metasearch service[17], under the 'content baskets' section. We have created a set of CRUD operations designed to be used from a web interface perspective, using them the orchestrator can modify or delete the content of one or several baskets or delete several of them at the same time.

[17] https://xreco.ari-imet.eu/api/docs

6 Search services

6.1 Local search backend (UNIBAS)

The search backend is an integral part of the *Neural Media Repository (NMR)* and therefore described in Section 3.3.

6.2 3D search demonstrator (CERTH)

We have developed a standalone User Interface (UI) to showcase our MuseHash approach in 3D mesh retrieval. This UI functions as a demonstration of the MuseHash method and will not be integrated into the final UI for the XReco project.

The XReco user interface (UI) offers a user-friendly platform for accessing and manipulating 3D models. At its core, it facilitates a 3D model retrieval engine that can be used to detect similar models inside the same dataset. The interface consists of three main components: the initial results panel, the similar results panel and the 3D model viewer.

The first element users encounter upon entering the UI is the initial results panel, as illustrated in Figure 6-1. When users click a button, the similar results panel (Figure 6-2) is displayed on the right, appearing next to the initial results panel, thus providing a dual-panel view. The 3D model viewer (Figure 6-3) appears as a popup modal when the corresponding button is clicked. These three components are described further below.

Upon entering the interface, users can observe the initial results panel that displays snapshots of the models. On the top of this component, there is a dropdown list from which users can navigate to any of the available datasets and view their respective models (i.e., ModelNet4.0, BuildNet_v0). A pagination control is also provided which ensures easy navigation, enabling transitions between pages that contain 100 models each. These two dropdown lists (pagination control and dataset dropdown) are shown and highlighted in red squares in Figure 6-1.

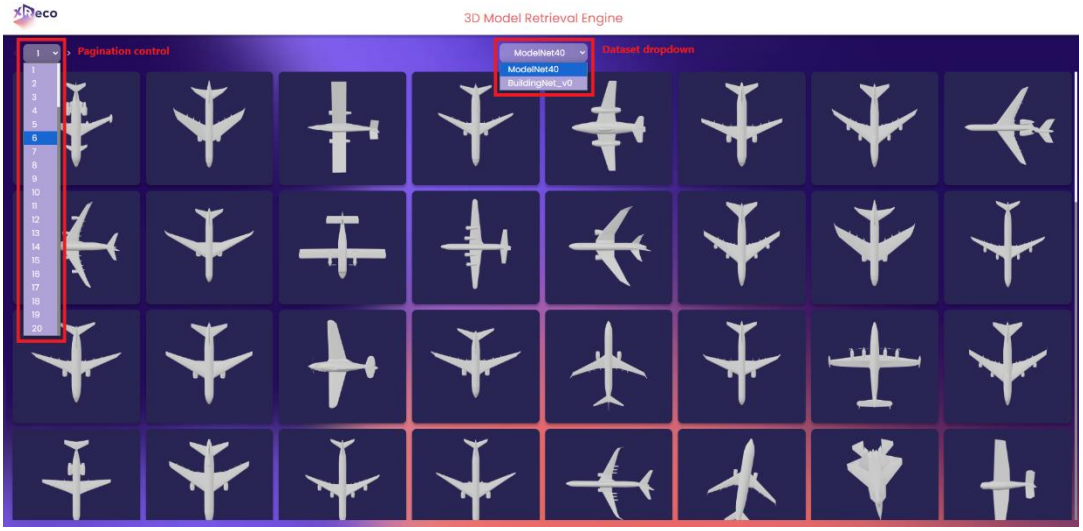


Figure 6-1 The initial results panel and its dropdown lists.

By hovering over a snapshot, a user can see the model's name, along with two buttons. The download button allows the user to select and download either the snapshot (in a PNG format) or the 3D model itself (in a GTLF

format). The left button triggers a query to the multimodal service using MuseHash (Section 4.4.2). After that, MuseHash returns models like the chosen one and opens the similar results panel. The highlighted model being hovered over is depicted in Figure 6-2, showcasing these details within a red square.

The similar results panel (Figure 6-2) displays the 50 more relevant models to the selected one. It should be noted that for each one of these models, there are the same buttons and functionalities as in the initial results panel. Moreover, at the top of the panel, there is the snapshot of the selected model along with its name and the associated category or categories.

Figure 6-2 The similar results panel and a model's details.

Upon clicking on the model snapshot, the 3D model viewer emerges, offering users an interactive space to manipulate the models that involves rotating and zooming into the model. Finally, a button is available for downloading the 3D model in GTLF format.

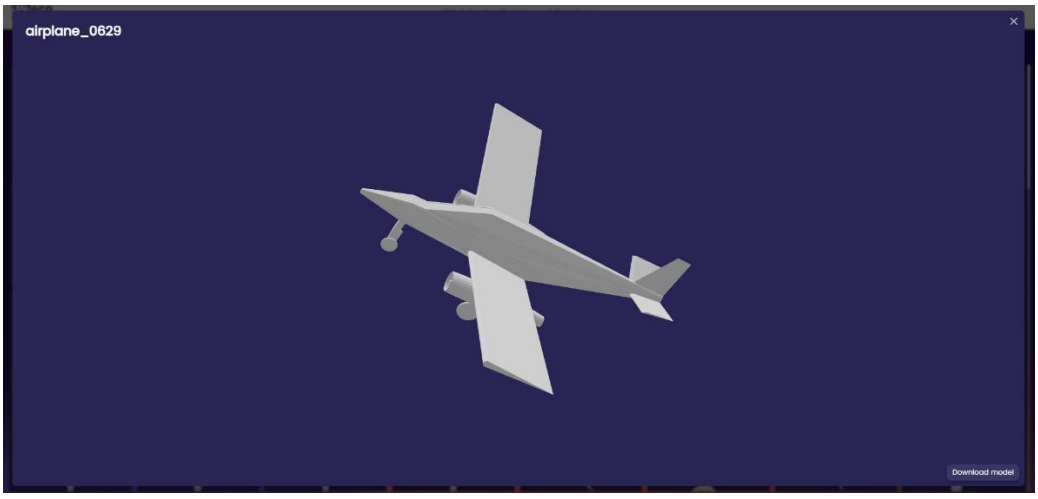


Figure 6-3 The 3D model viewer.

The interface is constructed using a technology stack that includes MongoDB as the database for storing model information, PHP for the backend, and HTML/CSS/JavaScript for the frontend. Google's model-viewer.js library is employed for the 3D model viewer.

Below we give two results for unimodal case and using only the mesh modality for ModelNet40 (Figure 6-4) and BuildingNet_v0 (Figure 6-5) dataset, respectively.

Figure 6-4 Results based on a specific mesh query from ModelNet40 dataset.

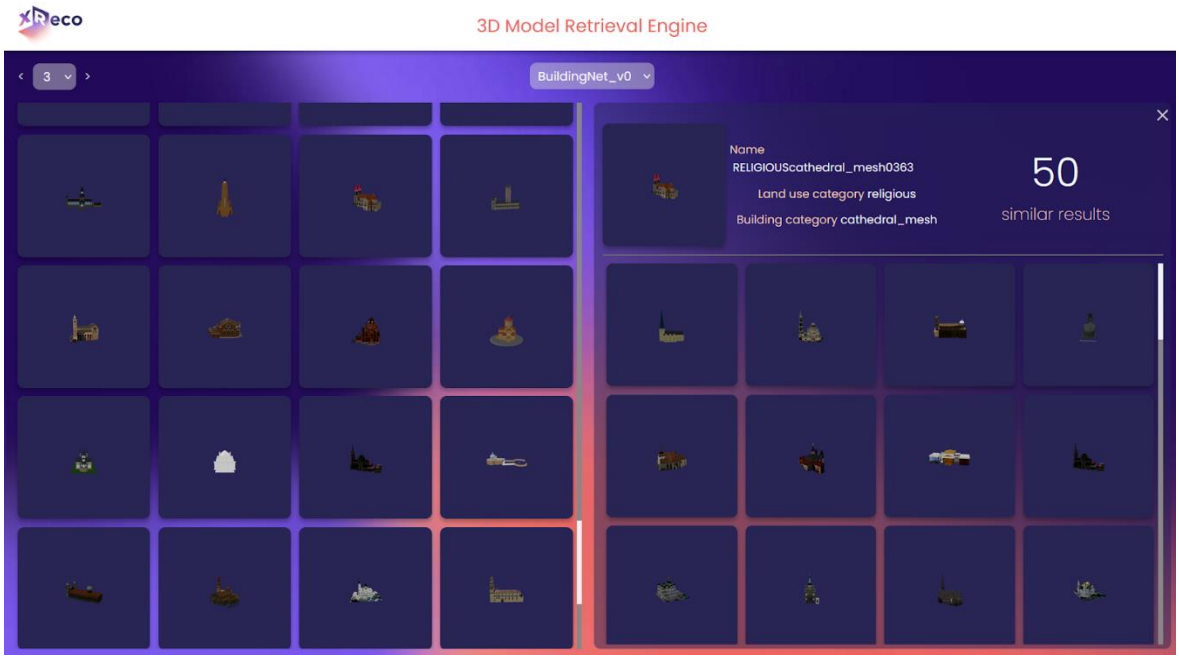


Figure 6-5 Results based on a specific mesh query from Buildingnet_v0 dataset.

6.3 Metasearch service (Atos)

A metasearch service is a type of search tool that aggregates results from multiple search engines or databases and presents them in a unified format. This definition aligns with T3.5, centred on developing a connection

between various repositories so that they are perceived by an end user as a single one. As described previously in connector part in Section 3, some of these repositories, such as those from RAI, DW, and UNIBAS, need to be internal to the project since they are provided by internal partners. We have also decided to include publicly available repositories to benefit from the vast amount of content they provide, and to gain a reference for the industry standard regarding this kind of endpoints. At present, we have successfully integrated all the partner provided repositories, the public repositories of Sketchfab and Wikimedia Commons into the first version of the tool. The ultimate goal is to develop a tool that is genuinely useful for the end user and easy to integrate with the other services of the project. Additionally, we aim to provide advanced functionalities that are appealing to creators and users of new types of media files. For instance, these functionalities could include filtering content by license type or by the different sources consulted. Additionally, there is currently no search engine that serves as the reference for 3D models on the internet. Therefore, after completing the refinement of the connectors and the development logic, we would like to add as many 3D repositories as possible to offer an updated state of the art tool.

6.3.1 Micro services Architecture and definition

The Figure 6-6 represents the current deployment of the metasearch service. As can be seen at first glance, the service is based on a microservices architecture. Only the components relevant to the metasearch will be described here, as the rest of the components are detailed in other sections of this document.

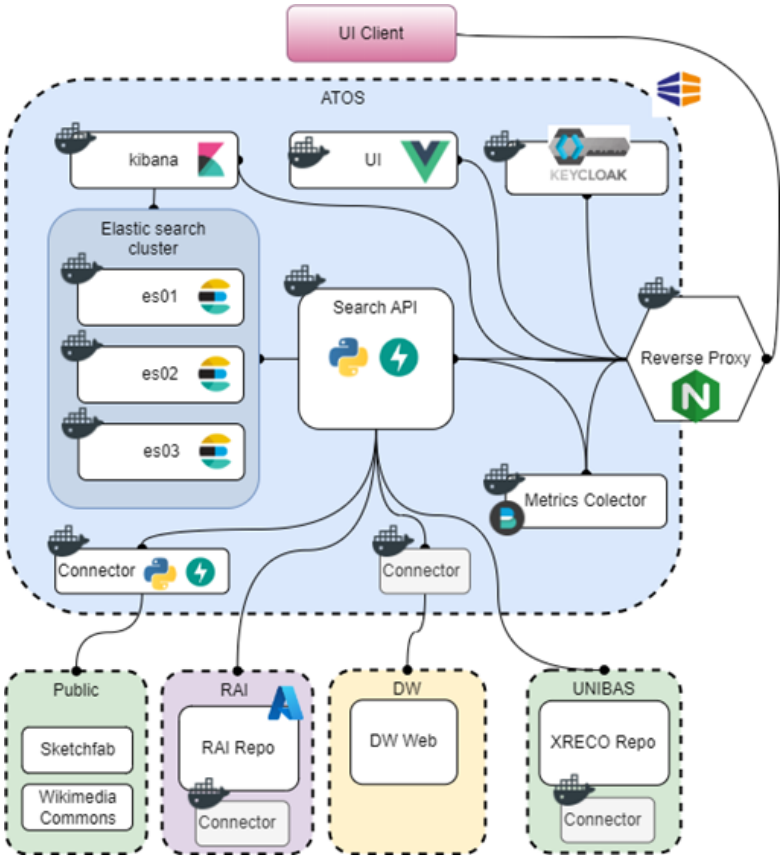


Figure 6-6 Metasearch Deployment

6.3.1.1 Search API

The search API container holds the functional core of the metasearch. It is essentially the component that: handles the initialization and connection with each of the connectors, executes all interaction with Elasticsearch, and launches the API and its public documentation. It is based on the FastAPI framework for Python. The logic of how the queries and results are ranked and cached will be explained in Section 6.3.2.

6.3.1.2 Elastic stack

We choose ELK[18] among the others we investigated for this purpose. Elastic offers a non-relational database that stores information in volatile memory and is capable of indexing as quickly as possible. Additionally, Elastic-stack offers indexing capabilities and compatibility with other systems widely used in the industry. From the Elastic Stack (ELK), we are using and customizing the following services:

- Elastic search cluster: The official recommendation from the Elasticsearch documentation is to create a cluster of at least three nodes for a production environment. This approach allows for sharding and load balancing, resulting in much faster response times for multiple users compared to using a single node. Additionally, having tested a deployment prepared for production gives us the capability to scale the development without too many complications.
- Kibana: Kibana is the official web interface for interacting with an Elasticsearch deployment. It provides the ability to perform troubleshooting more quickly and to represent data in a much more visual manner.
- Metrics Collector: We use a customized container with Filebeats to serve as a metrics collector, allowing us to create predefined dashboards with the information we receive from Nginx, or to create our own customized ones with the logs from the search API.

6.3.1.3 Auth and routing

The environment components include the reverse proxy, and the authorization and authentication services. The authorization and authentication are based on Keycloaks production deployment, which includes a container for Keycloak, and a PostgreSQL Database. The reverse proxy is based on Nginx, allowing us to conceal all services under a single url base, adding security by exposing only one port to the exterior, controlling all requests and managing the securization of the requests offering a unique SSL certificate. Additionally, we have integrated it with the metrics collector to monitor various parameters of the requests made against the metasearch, such as the country of origin, error control, or request frequency. We thought this information could be useful in later states of the development to further refine the ranking mechanism of the metasearch.

6.3.1.4 Metasearch UI

To demonstrate the progress of our development to end-users and provide an integration point with other partners in the consortium, we have provided a public endpoint since the first beta version of the service. Considering the optimization requirements of the metasearch, we have decided to use a modern, production-ready framework for the interface development, such as Vue.js.[19] The endpoint is publicly available at the following link: https://xreco.ari-imet.eu/search/. The Figure 6-7 shows the current appearance and available functionalities, including ranked connector results, pagination, filter by type, connector or License, result highlighting and redirection to the original source.

[18] https://www.elastic.co/es/elastic-stack
[19] https://vuejs.org/

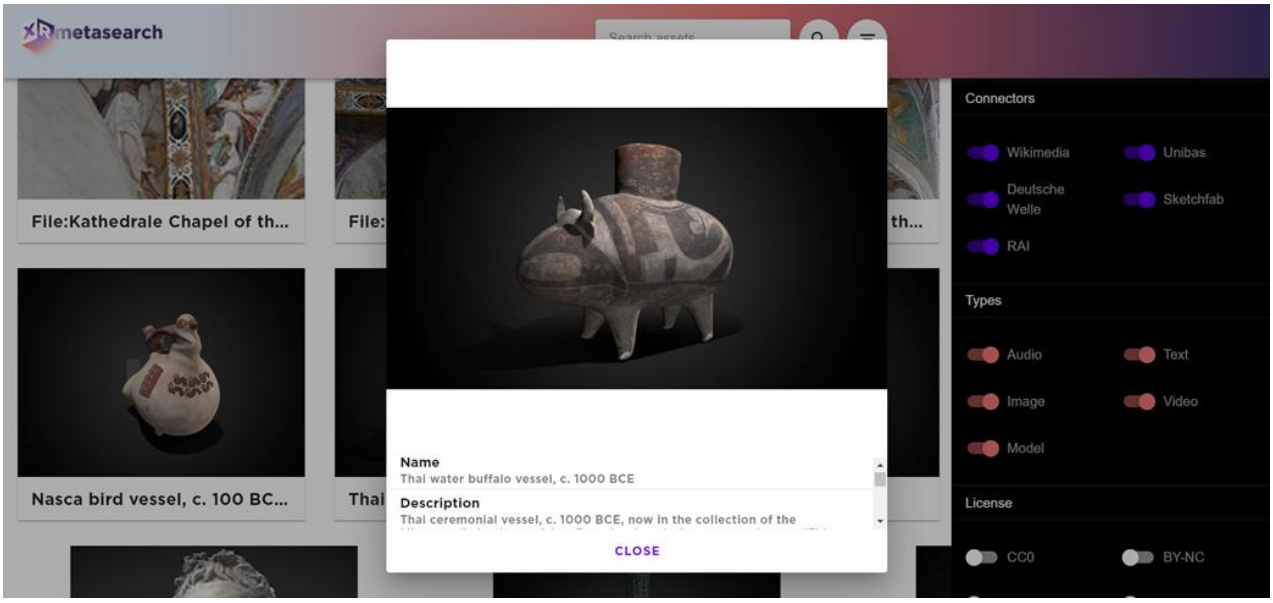


Figure 6-7 Metasearch User Interface

6.3.2 Metasearch logic

Our primary objective during the first part of this project has been to provide a solution that is integrated with the necessary partners and repositories and accessible to everyone. After achieving this objective, we aim to iterate on the service by leveraging our research lines and the feedback we may receive from other partners. Taking this into account, we have started with an effective solution that can easily merge the metadata from all connectors and has a reasonable response time.

We have created four indices in Elasticsearch to store the following information from the metasearch: the queries previously made to the service, the raw responses from the connectors, the ranked connector responses, and the content baskets. The basic logic of the service is depicted in the (Figure 6-8) flowchart. Essentially, we have developed a mechanism for ranking and caching responses.

Regarding caching, since we are going to consult the APIs of many connectors and receive many results, the process can be very time and hardware consuming. We needed a mechanism to respond faster to queries. Since the project focuses on European buildings of historical significance, we conducted an analysis of their search frequency on Google to assess the viability of this idea. This number reaches tens of thousands of searches per month for the most popular ones in each country. Our idea is to save on the consumption produced by all these queries and, additionally, to pre-populate the service with a dictionary listing all possible monuments.

The ranking mechanism, for now, is unimodal and text-based, creating a weight assignment within the connector. First, according to the relevance of each response's metadata attribute, and second, according to a weight for that specific attribute for each connector. In addition to this weight assignment, we are working on creating a background agent that examines cached results and influences the relevance of a specific result. For this, we are testing various artificial intelligence models that leverage the non-text-based information available to us, such as excerpts from the original content, thumbnails, or the content itself. We also plan to benefit from the services provided from other partners after the first integrated version of the platform for the same improved ranking mechanism.

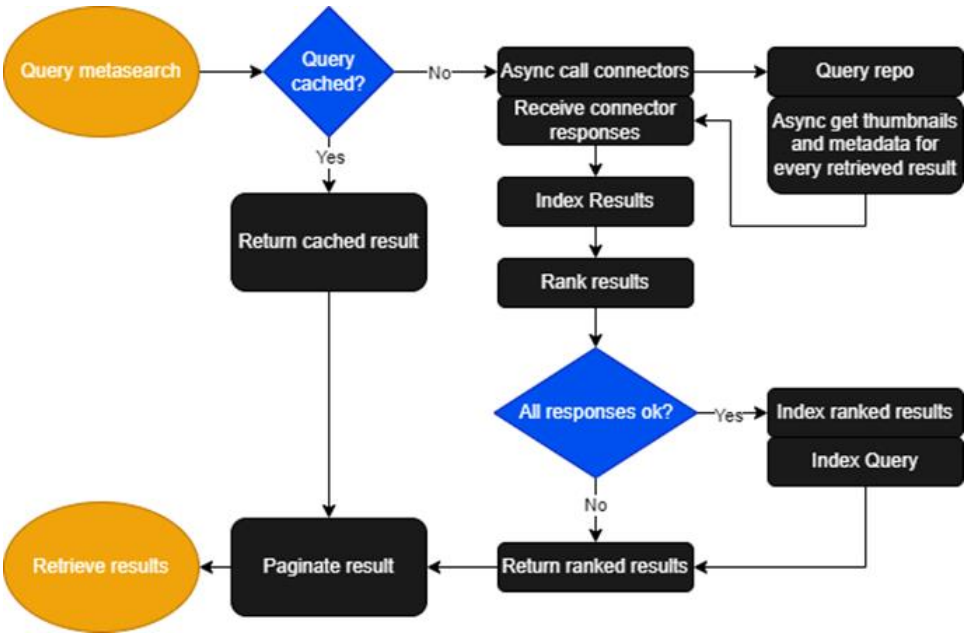


Figure 6-8 Metasearch flow chart

7 Legal Requirements for Rights Management

7.1 Introduction

Legal requirements are an important aspect of the research activities of the XReco project. The following positions the legal requirements section of the deliverable within the project, including its objectives, and indicates the relevant legal frameworks that are addressed in the subsections.

7.1.1 Context, Objectives and Positioning of the Legal Requirements Section

XReco is a project situated within a fast-moving technological, policy and regulatory environment. Through its focus on extended reality (XR) applications, it not only touches upon current developments in the context of "Web 4.0", virtual worlds and the "metaverse"[20], but it also raises novel questions regarding broader debates concerning digital phenomena, the EU's approach to the data economy, and established and emerging issues regarding intellectual property, especially copyright and related rights.

7.1.1.1 EU Policy and XR

The European Commission has developed a concrete **Web 4.0 strategy initiative** to ensure that the EU stays ahead of the curve regarding recent market developments, outlining key action points that aim to make sure Web 4.0 and virtual worlds, as part thereto, "reflect EU values and principles and fundamental rights, where people can be safe, confident and empowered, where people's **rights as** users, consumers, workers or **creators** are respected, and where European businesses can develop world-leading applications, scale up and grow" [21].

This policy intervention regarding Web 4.0 can be considered within a broader context of the **EU's approach for digital policymaking and regulation**, which is composed, among others, by the regulation of control and sharing of data at the EU level. Previous legislative periods have marked a new era for the data protection and privacy[22], interventions in the area of platform regulation[23], whereas the current legislative period has seen an increased assertiveness in addressing infrastructural, including standardization and interoperability, challenges[24]. This is a concrete goal of new legislative instruments such as the **Data Governance Act**[25] and **the Data Act**[26] and is also addressed by the EC's Web 4.0 initiative[27]. More specifically, the EU is set to establish a new framework for the

[20] Commission, 'An EU initiative on Web 4.0 and virtual worlds: a head start in the next technological transition' COM(2023) 442final, accessible at: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52023DC0442>(EC Web 4.0 Initiative)

[21] EC Web 4.0 Initiative, 4; the European Commission also launched a Citizens' Panel on virtual worlds to include the perspective of citizens in its policy on Web 4.0 (European Commission, 'Staff Working Document: Citizens' Panel Report on Virtual Worlds | Shaping Europe's Digital Future' (5 July 2023) <https://digital-strategy.ec.europa.eu/en/library/staff-working-document-citizens-panel-report-virtual-worlds>).

[22] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1.

[23] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance) [2022] OJ L 277/1 (DSA) and Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) [2022] OJ L 265/1 (DMA).

[24] Commission, 'A European strategy for data' COM(2020) 66 final (Data Strategy).

[25] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act, DGA) [2022] OJ L 152/1.

[26] Regulation (EU) 2023/ 2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) [2023] OJ L 71/1.

[27] EC Web 4.0 Initiative, 4.

exchange of data across and within specified sectors using the rubric of "**common European data spaces**"[28]. In the context of XR, the planned common European data space for media addressing "European publishers, broadcasters, radios, advertising companies, media SMEs, technology providers, content and tech start-ups, content creators and producers"[29], as well as for cultural heritage, which will "[boost] the digitisation of cultural assets and the reuse of high-quality data in this sector and others, such as media and tourism"[30], are set to be particularly important frameworks for enabling the exchange of XR-relevant data and content. Data spaces infrastructure, and its ongoing development in multiple fields,[31] may be an important aspect of the XReco architecture[32].

XR is also a unique issue for intellectual property, especially for copyright and related rights. This is also true for XR content that exists digitally. Therefore, effective and compliant **rights management** will be a key concern in XR content creation and sharing. Especially with the increased dissemination of artificial intelligence (AI) tools capable of generating or assisting in the creation of content, the European Commission's 2020 IP Action Plan makes clear that there remains room for improvement in the regulatory framework[33]. Within the context of mainstreamed Generative AI and growing concerns addressing issues of AI-generated content[34], this also raises questions concerning the **production of XR content facilitated by AI tools**. The use of such may also require an awareness of the developing environment of AI regulation, such as the EU's AI Act, including how it interacts with the application of relevant copyright law framework. XReco, therefore, directly contributes to the dynamic dialogue on aspects of IP, as further "realities" are unlocked, in the way forward to the development of Web 4.0 and virtual worlds/metaverse.

7.1.1.2 Positioning of this Section of the Deliverable

This section of the deliverable follows a legal research approach, focusing on **mapping and analysing potential legal issues and questions relevant for XReco in regard to the management and monetization of IP rights and data,** in accordance with the objectives of the project. Key objectives of XReco have legal implications, either by

[28]Data Strategy, 12; These data spaces are understood to "[bring] together relevant data infrastructures and governance frameworks in order to facilitate data pooling and sharing" (Commission, 'Commission Staff Working Document on Common European Data Spaces' SWD(2022) 45 final (Data Spaces SWD), 2); data spaces will include: "(i) the deployment of data sharing tools and services for the pooling, processing and sharing of data by an open number of organisations, as well as the federation of energy-efficient and trustworthy cloud capacities and related services; (ii) data governance structures, compatible with relevant EU legislation, which determine, in a transparent and fair way, the rights of access to and processing of the data; (iii) improving the availability, quality and interoperability of data – both in domain-specific settings and across sectors.".

[29] Data Spaces SWD, 36.

[30] Data Spaces SWD, 38; see also: Commission Recommendation of 10.11.2021 on a common European data space for cultural heritage [2021] OJ C(2021) 7953 final.

[31]See for instance: DSSC, 'Community of Practice – Communities – Data Spaces Support Centre' (*DSSC*, 2023) accessible at: <https://dssc.eu/space/DC/27983886/Community+of+Practice>.

[32] See Task 3.5.

[33] Commission, 'Making the most of the EU's innovative potentialAn intellectual property action plan to support the EU's recovery and resilience' COM(2020) 760 final, 2.

[34]Stuart A Thompson, 'A.I.-Generated Content Discovered on News Sites, Content Farms and Product Reviews' *The New York Times* (19 May 2023) <https://www.nytimes.com/2023/05/19/technology/ai-generated-content-discovered-on-news-sites-content-farms-and-product-reviews.html>; The Economist, 'Now AI Can Write, Sing and Act, Is It Still Possible to Be a Star?' *The Economist* <https://www.economist.com/briefing/2023/11/09/now-ai-can-write-sing-and-act-is-it-still-possible-to-be-a-star>; Laurent Carpentier, 'AI and Culture, Friends or Foes?' *Le Monde.fr* (7 October 2023) <https://www.lemonde.fr/en/culture/article/2023/10/07/ai-and-culture-friends-or-foes_6154693_30.html> accessed 11 December 2023; this has also received attention from copyright scholars: Theodoros Chiou, 'Copyright Lessons on Machine Learning: What Impact on Algorithmic Art?' (2020) 10 JIPITEC <https://www.jipitec.eu/issues/jipitec-10-3-2019/5025>; João Pedro Quintais, 'A Primer and FAQ on Copyright Law and Generative AI for News Media' (*Medium*, 26 April 2023) <https://generative-ai-newsroom.com/a-primer-and-faq-on-copyright-law-and-generative-ai-for-news-media-f1349f514883> accessed 11 December 2023.

being directly linked to licensing or monetization, or by addressing content search, analysis and potential reuse in the context of creation and dissemination of 3D content. This section of the deliverable contributes toward the key purpose of providing "the **concept of the rights management** and **licensing framework**" and a "**first version of licensing components"** of the XReco platform[35]. This section of the deliverable is therefore **legal in nature** – a key stepping stone in the state of legal research of aspects of the XReco extended reality data sharing platform, providing guidance, with a view to informing and empowering relevant stakeholders.

The technical components and aspects of the XReco workflow, as they are currently envisioned, are very important. They inform the identification, mapping and analysis of legal requirements for rights management and licensing framework. The monetization component will additionally be defined by legal requirements detected for rights management and licensing framework at a minimum grade and may be additionally crafted according to the business model vision.

To this end, **a)** it identifies and maps the ***legal requirements*** relevant to rights management, including licensing framework and monetization component and **b)** formulates ***legal analysis*** based on these requirements, indicating potential risks and opportunities for stakeholders that may emerge.

7.1.2 Methodology

Methodologically, this section of the deliverable employs legal desk research. Specifically, the method pursued is intended to determine the applicability of the relevant legal framework (see Section 7.2) to XR content production and sharing workflows. In order to achieve this, the research sets out via a description of (potential) XR content production and sharing workflows, providing a storyline of the XReco platform. This is descriptively contextualised by the relevant EU legal framework for the workflow in question, in due consideration of the scope and aims of the XReco project. Finally, this method involves an assessment of the workflow(s), based on the applicable relevant legal framework, indicating key issues and delivering a first version of licensing components.

In this context, the notion of "extended reality" is understood to encompass a range of technologies "which enhance reality and our senses by adding digital information to the real world or creating a new digital environment altogether. [36]" In light of this definition, the underlying notion of "digital information" is understood broadly, though it is recognised that overlaps with specific legal definitions may emerge[37].

7.1.2.1 Choice of EU law

A substantial amount of this section of the deliverable addresses legal issues and/or is based on legal resources. As XReco is a project addressing the needs and interests of a variety of stakeholders across the European Union (EU), this legal research focuses on the law of the EU. This does not prejudice the potential relevance of

[35]DoA, 35.

[36]EC Web 4.0 Initiative, 87.

[37] Specifically, "data" and "content". Data are generally the representation of facts. Data is defined by the Data Governance Act as "any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audiovisual recording" (Art. 2(1) DGA; Art. 2(1) Data Act); digital content is defined by the Digital Content Directive as "data which are produced and supplied in digital form" (Art. 2(1) Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services [2019] OJ L 136/1 (Digital Content Directive)); these legal definitions may not always be a perfect match for the technical processes described. For instance, whereas information represented digitally may be understood to comprise a type of data, "information" as such is not defined in EU law, and further may be subject to different rhetorical usage. In light of this, where these definitions are especially pertinent, this is addressed directly.

international legal sources of authority that may be common across EU Member States, nor the potential relevance of instances of applicable law of the Member States. The purpose of this scope is therefore for the research to be equally relevant for cross-EU stakeholders.

In light of this, it is important to clarify the relevant areas of law (Section 7.2) and Legally-Relevant Technical Aspects of XReco Workflow (Section 7.3), before proceeding with the legal assessment containing guidance over key legal aspects, risks & opportunities (Section 7.4).

7.2 Overview of Relevant Areas of EU Law for Legal Analysis

With EU law as an overarching framework, it is important to highlight **which areas of law within EU law are especially relevant for XReco as a data sharing and XR creative platform**. Data and content, including in the context of XR, may trigger special attention of certain areas of law, which are identified here. XReco as a platform might give rise to further questions about its legal governance[38]. Generally, EU laws[39] (also referred to as secondary legislation) may either be directly applicable across EU jurisdictions, or it may be in need of being transposed and implemented by the EU Member States in the national law. The former is the case for Regulations, whereas the latter is true for Directives[40].

7.2.1 Copyright Law Framework

Most crucially, "content" as a creative output of various forms implicates copyright and related rights issues. This is an area of law that requires careful attention in the potential management of rights and monetisation of "assets", as described in the DoA, and that, therefore, is the focus of discussion in this section of the deliverable. As any given digital file, "asset" or other piece of data or content may involve certain rights of authors or other rightsholders, the copyright framework is the most relevant[41]. For instance, a 3D model made from scratch by a talented artist may typically be understood to be protected under copyright, however, where said 3D model integrates certain other material (for instance, other 2D works), it may be more difficult to understand what can and cannot be done with the final 3D model. The purpose of this following discussion is therefore also to clarify situations such as these.

Copyright law in the European Union is primarily addressed at the level of the national law of the Member States. Nevertheless, in light of the EU's various competences to legislate, there is a significant EU-level body of law that addresses copyright and related rights. In light of the EU-level nature of the XReco project, this EU-level of copyright and related rights legislation is the focus here[42].

7.2.2 Data & Platform Regulation

XReco aims to create a platform which can manage large quantities of data, some of which might include personal data[43]. Consequently, it is necessary to address the potential impact of XReco on the right to privacy and the protection of personal data. In this context, the relevant legal frameworks include, particularly, the

[38] It should be underlined that this is informed by the overarching vision for XReco, meaning that some areas of law (both public and private) that may in some way be connected to certain dynamics related to the project or its outputs are excluded from consideration.
[39] For a detailed overview of the relevant legal frameworks, including references to case law and relevant legal abbreviations, please consult the Annexes.
[40] Art. 288 Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C 326/47.
[41] "Data" or "content" may encompass both copyrighted works and elements not covered by copyright.
[42] For a broader overview of the copyright framework, please consult Annex VI.
[43] Personal data is defined as "any information relating to an identified or identifiable natural person ('data subject')" (Art. 4(1) GDPR).

European Convention on Human Rights (ECHR) [44], the Charter of Fundamental Rights of the European Union (CFR) [45], and the General Data Protection Regulation (GDPR) [46].

The GDPR specifically outlines rules "relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data" [47]. Privacy and personal data protection are connected legal categories, but subject to important distinctions.[48] If use cases or architecture do not involve the processing of personal data, data protection laws generally do not apply[49]. Similarly, if no information related to individuals is being processed, the right to privacy might not be implicated. Crucially, processing is defined as "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means" [50]. The most practical approach, in order to avoid challenges of compliance and minimize risks from unlawful processing of personal data for a project such as XReco is therefore not to process personal data.

Data is an increasingly important area of regulation in the EU beyond the area of personal data (see Annex V). Given that XReco platform is based on data-driven technologies, assessment of these frameworks is also relevant. In addition, platforms that use digital technologies have been regulated more recently so that rights and obligations can be realised in the context of new digital phenomena. In that regard, implications from the application of such frameworks in the XReco platform are also relevant (see Annex V).

7.2.3 AI Regulation

XReco also relies on the implementation of certain AI technologies. In April 2021, the European Commission proposed a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [51]. At the time of writing, this Regulation is undergoing finalisation in the legislative process. These include rules on the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') [52], the prohibition of certain artificial intelligence practices[53], specific requirements for AI systems deemed high-risk and

[44]Art. 8 Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR).

[45] Art. 7Charter of Fundamental Rights of the European Union [2016] OJ C 202/389 (CFR).

[46] The ECHR and the CFR are instruments of international law, potentially subject to national implementation, whereas the GDPR is a Regulation of the EU, meaning that it is directly applicable across all EU Member States.

[47]Art. 1(1) GDPR. Cf. Artt. 2 and 3 GDPR regarding the material and territorial scopes respectively.

[48]Juliane Kokott and Christoph Sobotta, 'The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3 International Data Privacy Law 222, 225; Privacy has a broad scope and includes aspects such as "private and family life, home and communication" (Art. 7 Charter of Fundamental Rights of the European Union); data protection specifically focuses on the handling and processing of personal data. While privacy may cover a wider range of information, data protection is more specific. While privacy and data protection can overlap, there can be important instances where they diverge.

[49] Though users of anonymized data should be aware of re-identification risks (see Michèle Finck and Frank Pallas, 'They Who Must Not Be Identified—Distinguishing Personal from Non-Personal Data under the GDPR' (2020) 10 International Data Privacy Law 11).

[50] Art. 4(2)(GDPR.

[51]AI Act Proposal.

[52]Art. 1(a) AI Act Proposal; defined as "software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" (Art. 3(1) AI Act Proposal); Annex I: '(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods." For a more analytical approach on AI definitions, see e.g. Sofia Samoili et al, 'AI WATCH. Defining Artificial Intelligence', (2020) EUR 30117 EN, Publications Office of the European Union, available at: <https://publications.jrc.ec.europa.eu/repository/handle/JRC118163?mode=full>, *passim*. Cf. the distinction between "fully-generative machines" and "partially generative machines" at Jane Ginsburg and Luke Ali Budiardjo, 'Authors and Machines', 34 (2019) Berkeley Tech L J 343, 411ff.

[53]Art. 1(a)[sic] AI Act Proposal.

obligations for operators of such systems[54], harmonised transparency rules for AI systems intended to interact with natural persons, emotion recognition systems and biometric categorisation systems, and AI systems used to generate or manipulate image, audio or video content[55], and rules on market monitoring and surveillance[56].

The reach of this regulation is potentially vast, with certainty regarding key questions still being unresolved. For all such AI systems, providers and users can be expected to face the following regulatory interventions:

- Being subject to transparency obligations[57];
- A new regime for post-market monitoring, information sharing and market surveillance[58];
- Encouragement to draw up codes of conduct in regard to AI systems[59];
- A requirement for providers of foundation models[60] to "document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law" [61].

XReco's XR Services, as described below, even if they are not fully autonomous, may fall under relevant definitions and may qualify as AI systems **implementing generative AI tools**. Accordingly, their output (**New 3D assets**) could be **qualified as AI-generated content** that is produced by means of user's interaction with the XR service. User's interaction with the AI system mainly involves the definition of input data (ingested content) but there mihgt be no user involvement at the output generation process.

7.3 Overview of Legally-Relevant Technical Aspects of the Workflow

This section outlines the workflow currently envisioned by the project, which will serve as the basis for a preliminary legal analysis of the rights management and licensing components.

This overview has the purpose of providing insights into which real-world XReco data and content scenarios can already be identified. This overview therefore not only provides a benchmark for understanding which legal challenges and opportunities can emerge from the practical design of XReco scenarios, but also to further make the legal analysis comprehensible. The motivation behind this overview is to provide a closer picture of the data and content at stake as well as their processing, so that the legal landscape for the specificities of XReco can be more accurately assessed and that a preliminary legal guidance may be provided.

7.3.1 XReco Workflow

In general terms, potential options for the XReco workflow involve at least four instances:

(1) **ingestion** of pre-existing two-dimensional (2D) content and data ("parent asset");

[54]Art. 1(b) AI Act Proposal.
[55]Art. 1(c) AI Act Proposal.
[56]Art. 1(d) AI Act Proposal.
[57]Art. 52 AI Act Proposal.
[58]Artt. 62-68 AI Act Proposal.
[59]Art. 69 AI Act Proposal.
[60]Defined as "Providers of foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("generative AI") and providers who specialise a foundation model into a generative AI system" (Art. 28b(4) Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (AI Act EP Draft), accessible at <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html>).
[61]Art. 28b(4)(c)AI Act EP Draft.

(2) **search and retrieval** of pre-existing two-dimensional (2D) content and data;

(3) **extended reality (XR) services** for the reconstruction of three-dimensional (3D) assets based on ingested 2D content as per (1), and;

(4) a **marketplace** for extended reality (XR) three-dimensional (3D) assets, either reconstructed as per (3) or pre-existing.

Ingestion: Ingestion is understood here as the sourcing of content and data, including from repositories or collections, but also individual pieces of content, entailing the ability to access that content and data through what is called a "neural media repository" (NMR), i.e. a database storing metadata about content and references to one or more version of the media file in a local repository. In this instance, access corresponds to the passing-through of e.g., uniform resource locators (URLs).

These content and data sources are **decentralised** and may be publicly available or in direct connection with XReco. While search across the federated repositories is possible using the metadata provided by the data sources, only ingested content could be then subject to search and retrieval that makes use of additional content analysis and descriptor extraction.

Search and retrieval: Search and retrieval is a core component of the XReco platform, which enables finding and, where relevant, the use of suitable content and data in the context of XR services. Search and retrieval is intended to be centrally hosted, though the exact nature of hosting at time of writing is not finalised.

The technical dimension of search and retrieval is likely to include the following components:[62]

- A neural media repository backend, storing metadata about content;
- Temporal structure and keyframes, extracting temporal structure metadata from content;
- Classification and detection, assigning labels to images, image regions or moving regions in content;
- The training of a classification and detection model;
- Multimodal descriptor extraction and matching comprising the extraction of a set of descriptors that can be used for similarity search;
- Metasearch sending requests to search interfaces, obtaining metadata and references to preview items.

XR Services: Some of the XR services developed in XReco involve chiefly technologies that enable the reconstruction of an object in three dimensions, including tools that allow for further manipulation and authoring, i.e., the tools needed to create novel media content with an XR orientation (see D4.1). XR services are intended to be centrally hosted, though the exact nature of hosting at time of writing is not finalised.

The technical dimension of XR services is likely to include the following components:[63]

- NeRF algorithms;
- SfM Reconstruction;
- Super resolution;
- Face mesh building.

Specific forms of 3D reconstruction that are envisioned by the XReco platform are outlined as follows:

[62] Based on information provided by the project partners.
[63] Based on information provided by the project partners.

- **Instant NGP (NeRF):** During the operation of this service, ingested content (i.e., a set of images of the same scene, along with their camera poses, or a video) is transmitted from the original source (i.e. copy in the NMR, or remote content item resulting from data mining) to the server (physical machine) hosting the Instant NGP service and is temporarily stored on the server's local disk. During algorithm training, whose features may vary, depending on the type of ingested content (images with known camera poses, images with unknown camera poses and videos), the content is loaded into the server's CPU RAM, and from CPU RAM to GPU RAM for faster/parallel access. When training ends, the ingested content is deleted.
- **NeRF+Depth:** During the operation of this service, ingested content (i.e., a set of images of **the same scene** taken by several camera sensors of the same model, from different positions and with different viewpoints, along with their camera poses or a video[64] is transmitted from the original source (i.e. copy in the NMR, or remote content item resulting from data mining) to the server (physical machine) hosting the Instant NGP service and is temporarily stored on the server's local disk. During algorithm training, whose features may vary, depending on the type of ingested content (images with known camera poses, images with unknown camera poses and videos), the content is loaded into the server's CPU RAM, and from CPU RAM to GPU RAM for faster/parallel access. When training ends, the ingested content is deleted.
- **NeRF in the wild:** During the operation of this service, ingested content (i.e., a set of images of **the same scene in different illumination settings/(foreground-background conditions)** along with their camera poses or a video) is transmitted from the original (i.e. copy in the NMR, or remote content item resulting from data mining) to the server (physical machine) hosting the Instant NGP service and is temporarily stored on the server's local disk. During algorithm training, whose features may vary, depending on the type of ingested content (images with known camera poses, images with unknown camera poses), the content is loaded into the server's CPU RAM, and from CPU RAM to GPU RAM for faster/parallel access. When training ends, the ingested content is deleted.

Marketplace: The marketplace for 3D XR assets is understood as a facility that enables the commercialisation of individual or multiple 3D content and content and data for their use in commercial applications in return for remuneration. A marketplace for XR assets may be federated.

The following legal analysis addresses chiefly data and content that moves through the four stages highlighted above. From a copyright law perspective, licensing agreements will be relevant only for copyright-sensitive acts of use/exploitation of protected material that are expected to take place during the XReco operation, according to the instances of workflow inventoried above and insofar an exception or limitation is not applicable. This assessment will offer an outline of the licensing needs, from which the first version of XReco licensing components may be extracted.

7.3.2 Assumptions

It should be noted, however, that in the XReco workflow there currently exist several options for the integration of separate technical components. Each variation may have important legal implications, with novel or different potential risks and opportunities than those assessed in the following. Looking ahead, a new assessment based on technical, business or legislative changes may be needed.

[64] In both cases, depth information (geometry representation) is also included.

The envisaged XReco storyline for workflow implementation relies on the generation of 3D data and content rooted in 2D data. Moreover, it comprises the use **of lawfully-accessible, authorised two-dimensional ingested data and content from sources that in fact provide such lawful access**. These data and content sources are decentralised and may be public-facing or in direct connection with the XReco platform.

Following that assumption, any content providers either own/have already licensed data and content for uses necessary for the operation of search and retrieval and XR services provided by XReco, or they provide lawful access to and further use of relevant data and content in the XReco context. This means that XReco would not facilitate the implementation of licenses or licensing terms for the ingestion of data and content. Content providers should have already cleared the relevant rights over ingested content[65]. Search and retrieval as well as XR services rely on the lawful access to and availability of authorisations to use the ingested data and content for their operation[66].

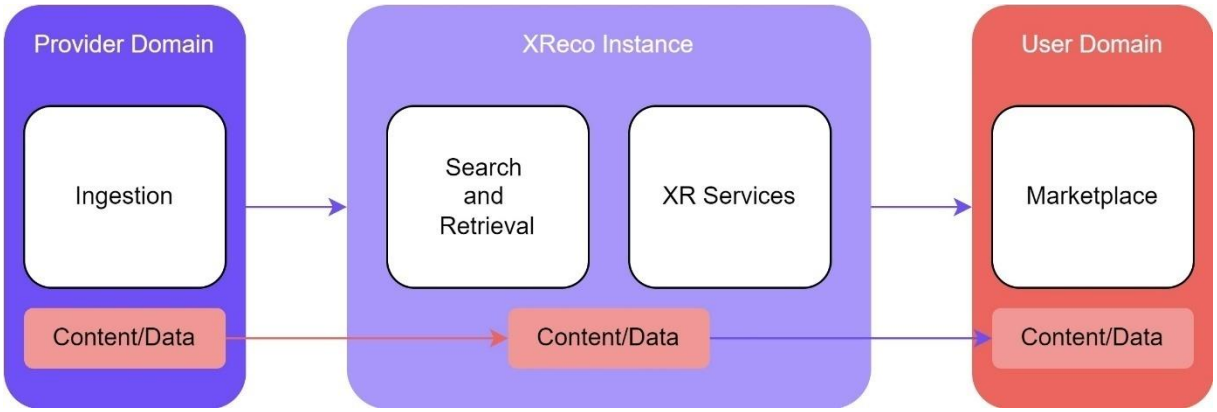


Figure 7-1 Simplified linear representation (NB: The chain of events may have iterative parts)

7.3.3 Disclaimer

It should be underlined that the material provided in this deliverable does not represent a legal endorsement, nor an assessment of the lawfulness of the practices envisioned or described[67]. Rather, it provides a benchmark for further legal research analysis that will follow. In this context, crucial questions cannot be addressed at this time, especially given the research purpose of this deliverable. These include, but are not limited to, the exact identity of the persons and/or institutions engaged in the particular practices, the minimum necessary rights and obligations pertaining to particular practices, exact conditions for uses of XR data and content. Importantly, a variety of media types can be input or output from the XReco ecosystem. While the focus rests on the generation of 3D data and content rooted in 2D data and content, other scenarios may also be possible within the framework of XReco. Finally, the form and content of this workflow will necessarily be addressed by dynamic changes to technical state of the art and business model considerations, which remain outside the scope of this section of the deliverable. Also, this deliverable describes the state of legal research about half way into the project. This research will continue and be adapted depending on the further development of the project as such. A final report on these research activities will be part of D3.2 (M32).

[65] This is discussed further in Section 7.4.1
[66] This is discussed further in Section 7.4.2
[67] The purpose of this deliverable is entirely informational and shall not be understood to comprise legal advice.

7.4 Analytical Legal Assessment: Key Legal Aspects and Risks & Opportunities

The following addresses each stage of the technical aspects of the XReco workflow, addressing and analysing key legal aspects, identifying risks and opportunities.

7.4.1 Data and Content Ingestion

7.4.1.1 Key Legal Aspects

7.4.1.1.1 Personal Data Protection Rules

Practically any handling of personal data would trigger the application of the GDPR and its rules. The GDPR outlines key principles that should be considered when processing personal data[68]. At the stage of ingestion, the most practical approach, in order to avoid challenges of compliance and minimize risks from unlawful processing of personal data for a project such as XReco is therefore **not to process personal data**. This can be achieved by the **use of anonymous information or data**, such that a potential data subject is not or is no longer identifiable via the data in question[69]. Anonymous information or data should be distinguished from "pseudonymisation" [70], which while recognised as reducing certain risks, does not preclude the application of data protection rules[71]. The GDPR provisions on personal data also do not apply to data about deceased persons[72], or AI-generated non-existent human faces, meaning that such data can be processed without additional burdens[73].

7.4.1.1.2 Copyright Exceptions and Limitations

In the context of data or content ingestion in some way, copyright exceptions and limitations may be an avenue of providing legal authorisation for the functioning of the ingestion process.

At the stage of ingestion of pre-existing 2D content and data, **the temporary reproduction and the TDM exceptions may be available**[74].

a) Where the five cumulative conditions of the temporary reproduction exception are fulfilled, lawful uses of the 2D content and data may be carried out[75].
b) Where the ingestion is carried out via an automated analytical technique, which is either carried out by a research organisation/cultural heritage organisation for purposes of scientific research or carried out without an opt-out by the relevant rightsholder being present, information may be generated from the data and content in question[76].

At the point of ingestion, in order to benefit from such exceptions, it may be necessary to ensure that the rights of the relevant rightsholder of the underlying object are not infringed[77], and/or that the digital image is not an

[68] Art. 5 GDPR.
[69] See Recital 26 GDPR.
[70] Defined as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person" (Art. 4(5) GDPR).
[71] Recital 28 GDPR.
[72] Recital 27 GDPR.
[73] Including e.g. for archival or historical research purposes (Recitals 158 and 159 GDPR).
[74] See Annex VIII regarding further detail on the TDM exceptions.
[75] See Annex VIII regarding the temporary reproduction exception.
[76] See Annex VIII regarding the text and data mining exceptions.
[77] This is an especially important consideration where commercial purposes are pursued, considering that such purposes may nevertheless be limited in light of the three-step test (see Annex VIII).

unlawful copy of the underlying object. Further, it may be important to verify that the data or content ingested is either not protected (for instance, where it is not an original work of authorship or if it is in the public domain), that it is licensed under a sufficiently-permissible license (e.g. CC-0[78]), or failing that, it is lawfully accessible. At the level of designing the means of ingesting data and content, this entails that the ingestion of content and data would be carefully vetted for each potential connector.

7.4.1.1.3 Copyright Licenses

At the stage of ingestion of pre-existing copyrighted works or other subject matter (2D protected content), and in the absence of application of an exception or limitation and in order to comply with applicable copyright legislation, the following copyright-sensitive acts need to be licensed:

7.4.1.1.3.1 Reproduction of works or other subject matter in the NMR backend repository ("upload license")

Ingestion of protected content **is triggered by a user**[79] **in the provider domain** of the XReco platform. The user:

a) either manually selects a **pre-existing** local media file, already stored in an external local storage, and uploads it to the storage of the NMR (along with basic metadata) [80];

b) or manually selects a **newly created item** (new asset), as result of the implementation of XR services (e.g., by 3D reconstruction), either directly from the service or via the user's local disk (i.e., ingestion is the same as in the previous case);

c) or manually selects the content to be ingested (one or more items) from a **metasearch results list** that involves also import of metadata attached to the retrieved (displayed) content[81]. In this scenario, Metasearch is part of the XReco operation (see below under (2)).

In all of the above ingestion scenarios, each ingested file that embodies a work (protected content, typically a 2D image, video or a video frame) **is copied to the storage of the NMR**. This copy is qualified as (digital) **reproduction** (creation of a copy) **of the protected content**, which is subject to the reproduction right[82], and, thus, subject to licensing. In this case of licensing transaction, the licensor should be the rightsholder(s) of reproduction right (pertaining to copyright and/or related rights) regarding the protected (ingested/uploaded) 2D content or a license holder of the **reproduction right** (in this case: sub-licensor) of that content. The (sub-)licensee should normally be the user (**uploader**) that manually uploads the content at the NMR backend repository (a) and b) ingestion scenarios). A license will not be necessary at this point insofar **the uploaded content is user's own content (rightholder = uploader)**. Still, it is possible that a user uploads third party content.

The mere upload of the work at the storage server may not comprise an act of making available to the public[83]. However, in case that this uploaded copy is accessible to a public, an act of making available will also be at stake[84].

[78] Creative Commons, 'CC0' (*Creative Commons*) available at: <https://creativecommons.org/public-domain/cc0/>.
[79] The user of the XReco platform does not always coincide with the user of protected content, given that a user may be rightsholder of uploaded own content. In the context of the diagram above in Figure 7-1, the "user" is acting within the provider domain.
[80] Please note that storage in this external local storage qualifies as reproduction according to Art. 2 Infosoc Directive but this reproduction is assumed to be lawful (i.e. licensed or covered by an exception or limitation), therefore licensing of this reproduction is not relevant for the purposes of this section.
[81] The content retrieved is equally stored in external local storage (local repository). This storage is again qualified as reproduction but this reproduction is assumed to be lawful, as in case of ingestion based on content uploaded by the user.
[82] Art. 2 InfoSoc Directive.
[83] Christophe Geiger et al, 'The Information Society Directive' in Irine Stamatoudi and Paul Torremans (eds), *EU Copyright Law. A Commentary* (2nd ed, Edward Elgar 2021), 295.
[84] Art. 3 InfoSoc Directive.

7.4.1.1.3.2 Display of works or other subject matter on Metasearch results (Metasearch display license) and further ingestion (upload license)

Metasearch (T3.5) enables ingestion based on search and retrieval of 2D content through various external (remote) local repositories (public or not), with the support of **connectors** that feed the metasearch with metadata and references for content, fetched from these external repositories[85]. At this point and prior to ingestion of content included in the results of the metasearch (namely, the manual selection of retrieved content by the user), **no copying** of the searched/retrieved content is performed (copying in the NMR backend storage is performed ***once the user selects*** retrieved content to be ingested). However, retrieved content (typically, 2D images or keyframes[86]) is typically **displayed on user screen**, as part of search results and is perceivable (accessed) by the user. Then, the preview of the retrieved content item may be downloaded by the user and stored to user's device.

Insofar the display of retrieved content in form of preview involves the display of the expression of pre-existing works or other subject matter to users of the Metasearch[87], then this **display in the results page possibly** constitutes an **act of communication/making available to the public[88]** and thus in principle is subject to licensing, to the extent that **a new public is reached**[89]. The new public corresponds to the users of the XReco metasearch component that may access searched and retrieved content from a place and at a time individually chosen. In case of the display of a keyframe of an audiovisual work (e.g., a video), an act of communication to the public of the work and making available of the film (first fixation) is equally expected to take place, insofar the keyframe reproduces partially elements of the protected visual expression of the work. Displayed content may equally be subject to transient copying (on-screen copies) for browsing and on-screen display purposes and thus trigger also application of reproduction right[90]. In addition, eventual **storage of downloaded copy** of the previewed content qualifies as reproduction (or partial reproduction, in case of a keyframe of audiovisual content), covered by Art. 2 Dir. 2001/29 which is, in principle, subject to license. More crucially, the content (one or more items) selected by the user to be ingested (including import of metadata attached to the retrieved (displayed) and selected content) will be copied to the storage of the NMR and, thus, subject to reproduction right, similar to other ingestion scenarios (see above under (1)).

The licensor should be the rightholder(s) or a license holder (i.e. sub-licensor) of the involved copyright and/or related rights (communication/making available to the public and/or reproduction) over the protected retrieved (displayed) and possibly ingested pre-existing 2D content. The (sub-)licensee(s) should be the provider of the search and retrieval (including metasearch display interface/search) services and/or the user thereof.

[85] In a more technical sense, Metasearch receives user's search requests and sends these requests to public / other organisations' search interfaces and then obtains metadata and references to preview/HQ the retrieved external content. The implementation of the metasearch component is based on handling of URLs of searched/retrieved content.

[86] Keyframes may be fetched from the remote location to display the search result, if such keyframe is referenced in the metadata of the retrieved content.

[87] Cf. M. Borghi/S. Karapapa, Non-display uses of copyright works: Google Books and beyond, *Queen Mary Journal of Intellectual Property*, Vol. 1 No. 1, April 2011, p21–52 and es 21.

[88] Art. 3 InfoSoc Directive.

[89] A public is a new public where it comprises "a public which was not taken into account by the authors of the protected works when they authorised their use by the communication to the original public" (Cases C-403/08 and C-429/08 *Football Association Premier League and Others* [2011] EU:C:2011:631, para. 197).

[90] Art. 2 InfoSoc Directive.

7.4.1.2 Analysis: Risks and Opportunities

7.4.1.2.1 Risk

At ingestion stage, input content may be necessarily copied for technical reasons. **The fact that the ingested content has been lawfully accessed (e.g. through a public API) does not automatically entail that the uploader holds reproduction rights over accessed content**.

Certain risks emerge from the specific requirements for relevant exceptions and limitations to copyright. The requirements of the mandatory temporary reproduction and TDM exceptions are reiterated here, and are discussed in greater detail in Annex VII.

In order for an act falling under the right of reproduction to benefit from the temporary reproduction exception, it must fulfil the following five requirements[91]:

- The act is **temporary**;
- The act is **transient** or **incidental**;
- The act is an **integral and essential part of a technological process**;
- The sole purpose of that technological process is either to **enable a transmission in a network** between third parties by an intermediary or to **enable a lawful use** of a work or protected subject matter, and;
- The act does **not** have any **independent economic significance**.

In order for an act falling under the rights (right of reproduction[92], right of extraction[93]) to benefit from the TDM exception for the purposes of scientific research, it must fulfil the following requirements[94]:

- The act comprises **text and data mining** of works and other subject matter;
- The act is carried out by a **research organisation** or a **cultural heritage institution**;
- The research organisation or cultural heritage institution has **lawful access** to the works and other subject matter;
- The act is carried out for the **purposes of scientific research**;
- Copies of work and other subject matter are **stored** with an **appropriate level of security**.

In order for an act falling under the rights (right of reproduction Art. 5(a) Database Directive; Art. 2 InfoSoc Directive; Art. 4(1)(a) Software Directive; Art. 15(1) CDSM Directive[95], right of extraction, right of adaptation[96]) to benefit from the general TDM exception, it must fulfil the following requirements[97]:

- The act is a **reproduction or an extraction** of works and other subject matter;
- The works and other subject matter are **lawfully accessible**;
- The act is carried out for the **purposes of text and data mining**;

91 Art. 5(1) InfoSoc Directive.

92 Art. 5(a) Database Directive; Art. 2 InfoSoc Directive; Art. 15(1) CDSM Directive.
93 Art. 7(1) Database Directive.

94 Art. 3 CDSM Directive.
95 Art. 5(a) Database Directive; Art. 2 InfoSoc Directive; Art. 4(1)(a) Software Directive; Art. 15(1) CDSM Directive.

96 Art. 4(1)(b) Software Directive.
97 Art. 4 CDSM Directive.

- The use of the works and other subject matter **has not been expressly reserved** by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online (no "**opt-out**").

In order for technically-necessary copying to be covered by the **temporary reproduction exception, the fulfilment of the five requirements is essential** where such copying by the relevant user is not permitted (either by ownership or by license). Certain technologies and methods employed for ingestion may not fulfil these requirements.

In order for copying in the context of data collection at the ingestion stage to be covered by the text and data mining exceptions, the **fulfilment of their respective requirements is essential**. Only research organisations and cultural heritage institutions for the purposes of scientific research may benefit from the exception where an opt-out is expressed. All other cases of text and data mining must respect an opt-out in machine-readable form for the purposes of text and data mining. In every case, **access to the data or content** subject to text and data mining needs to be **lawful**.

In case of user upload, the user that triggers the ingestion **needs to hold respective reproduction rights** (as owner or as licensee)[98]. Also, the uploader should confer to the XReco service provider respective rights by means of further (sub)licensing, in case that additional reproductions (than the copy at the NMR storage) take place under the initiative of the XReco service provider, such as in case of migration of stored content at another location. Import of metadata may involve database extraction, covered by the *sui generis* right (Art. 7 para. 2(a) Database Directive).

Ingestion by implementation of **Metasearch involves additional licensing needs/challenges**, insofar the displayed (retrieved) and possibly ingested/downloaded content is not user's own content. Accordingly, ingested content from external sources in Metasearch ingestion scenario (e.g. from public or non-public APIs) should in principle be associated with a copyright license covering (to begin with) ingestion stage (and beyond).

All necessary above-mentioned licenses need to be granted **prior** to the realisation of any copyright-sensitive act and must also cover all protectable elements into the licensed content.

7.4.1.2.2 Opportunities

Exceptions and limitations to copyright **may address certain content** at the ingestion stage. This is highly dependent upon the content in question, the mode of ingestion including how the content is accessed, as well as the person claiming to benefit from the content. For (parts of) the ingestion stage to benefit from the exceptions and limitations, especially the temporary reproduction and text and data mining exceptions, the ingestion stage needs to be **carefully designed**, **carried out with caution** and be **responsive to factual changes**, such as the availability of the opt-out for text and data mining.

An input license associated with ingested content could anticipate and **cover all possibly involved rights** and, therefore, all relevant licensing needs for any ingestion scenario. At the user upload scenario, this license may be granted by the user by means of wrap-up license (terms & conditions acceptance). All these licensing challenges are usually resolved (at least, prima facie) when content is subject **to an open license**[99]. Alternative models for authorisations may emerge as discussion concerning copyright and AI develop[100].

[98] The holding of further rights may be required (see Annex VI).

[99] On a case-by-case basis, this can also be problematic, see: Martin Kretschmer et al, 'Copyright law, and the lifecycle of machine learning models' (2024, forthcoming) 55 IIC, 16f.

[100] Cf. Martin Senftleben, 'Generative AI and Author Remuneration' (2023) 54 IIC 1535.

Key Points

- Ingested data and content may contain personal data. In order to avoid issues concerning data protection and privacy, it is best not to access or use data and content containing living persons.
- Ingestion of protected content typically comprises one or more acts of reproduction (including for purposes of NMR).
- A differentiation in licensing needs exists between Metasearch ingestion and user upload ingestion.
- Temporary reproductions may be permitted at the ingestion stage without a license, provided that the legal requirements for such reproductions are met.
- Reproductions within the context of text and data mining may be permitted at the ingestion stage, provided that access to the content or data is lawful, and the specific further requirements are met.
- All necessary licenses need to be granted **prior** to ingestion/display and must also cover all protectable elements in the licensed content.
- Open content constitutes a promising alternative to rights clearance via individual licensing of (retrieved and) ingested content. Alternative models may yet emerge in the dynamic AI legal space.

7.4.2 Search and Retrieval

Beyond the acts that are connected with the display of search results that are discussed above, several types of computational analysis **over the copy of ingested content stored at the NMR** storage (see above under Section 7.3.1) may take place during the implementation of search and retrieval components (WP3 Services-pre-production stage). These components refer to[101]:

(a) temporal structure and keyframes extraction: scanning through the entire media item and storage of lo-res keyframes;

(b) Classification, detection (+ tracking) (pretrained only): analysis of the entire media file, creation of numeric and textual metadata describing the item;

(c) Classification, detection (+ tracking) (fine-tuned/custom): analysis of the entire media file for training purposes, keeping a list of the IDs of the content used;

(d) Multimodal descriptor extraction and matching: analysis of the entire media file, extraction of feature vector/set of feature vectors per media item.

7.4.2.1 Key Legal Aspects

7.4.2.1.1 Copyright: Exceptions and Limitations

At the stage of search and retrieval of pre-existing 2D content and data, that is, where search and retrieval tools are enabled concerning content that is ingested in some fashion, these need to be in compliance with the lawful uses that are permitted by the relevant exceptions/limitations.

Where the ingested content is not downloaded and/or re-hosted, but merely lawfully indexed or referenced, the requirements of the temporary reproduction exception may nevertheless need to be considered and complied with, depending on whether or not which parts of the search and retrieval tools are integral and essential to

[101] (a), (b) and (c) are researched in T3.1, whereas (d) is researched in T3.2.

their technological processes. In a general sense, the temporary reproduction exception would not be available where a given tool that forms part of the package of search and retrieval is not integral or essential to the technological process.

7.4.2.1.2 Copyright: Licenses – Reproduction for further analysis of ingested content (WP3 services) (Description Services License)

(Computational) analysis of each 2D content media file that takes place when implementing WP3 services components involves the creation of **transient copies** of the protected content embodied in that file during analysis, since the item needs to be decoded and represented in RAM of the processing machine. Moreover, low-resolution keyframes are expected to be **stored in the NMR backend** repository for as long as the initial copy of the content is in the repository.

All those acts correspond to transient/temporary and/or partial copies of the media file **and are qualified as reproductions** (creations of a copy) of the protected content, which are subject to the reproduction right,[102] and, thus, in principle, **subject to licensing.**

Licensor should be the rightsholder(s) of copyright and/or related rights over the protected (analysed) content or a license holder of the **reproduction right** (in this case: sub-licensor). The (sub)licensee in this case should be the service provider.

On the contrary, the extraction of features and the creation of numeric and textual metadata *per se*, such as feature vectors, **is not qualified as reproduction** of a work or other subject matter and is not subject to copyright licensing.

7.4.2.2 Analysis: Risks and Opportunities

Risks: At search and retrieval stage, analysed content needs to be copied for technical reasons, mainly on **a transient or temporary fashion**. Still, the reproduction right[103] over ingested content is triggered and the XReco service provider that undertakes those acts (even if triggered by the user) needs to hold reproduction rights over ingested and analysed content, being noted that these reproductions are separate reproduction acts from initial ingestion copy and that the reproduction made for ingestion purposes is expected to be made by the uploader (and not the service provider).

Opportunities: Where search and retrieval is concerned with purely factual elements of underlying 2D content or data, this will typically not trigger any copyright concerns[104].

Where temporary reproductions of underlying 2D content or data is made, including where such content is protected by copyright, the temporary reproduction exception may be available. This requires the fulfilment of the five conditions of the exception. Where automated analytical techniques are used for the generation of information at the Search and Retrieval stage, the text and data mining exceptions may also be available, where their respective conditions are fulfilled.

Of course, all licensing needs at this stage may be (explicitly) addressed by means of **one license, which may be also the input license (see above)** to the benefit of the XReco provider. Last, all these licensing challenges are usually solved (at least, prima facie) when content is subject **to an open license**.

[102] Art. 2 InfoSoc Directive.
[103] Art. 2 InfoSoc Directive.
[104] See Annex 2 regarding the idea-expression dichotomy.

Key Points

- Search and retrieval of purely factual elements that are not original expressions do not trigger any copyright concerns.
- The implementation of certain WP3 services may involve one or more acts of reproduction that need to be previously licensed, unless if exceptions or limitations apply.
- Where factual elements for the purposes of search and retrieval are extracted from protected content or data, the temporary reproduction or text and data mining exceptions may be available, subject to the fulfilment of their respective conditions.
- These acts are expected to be performed by the XReco service provider and not the user.
- Open content constitutes a promising alternative to rights clearance via individual licensing of analysed content.

7.4.3 XR Services

7.4.3.1 Key Legal Aspects

7.4.3.1.1 AI Regulation

At the time of writing, a political agreement has been reached on the final text of the AI Act[105], however, the agreed text has not been published. In the context of XReco, and in the absence of legal certainty in the exact content of the AI Act, which is not enacted yet, the current approach to follow the High Level Expert Group on Trustworthy AI's Ethical Guidelines is a sensible decision (see D1.3).

7.4.3.1.2 Copyright: Exceptions and Limitations

The provision of XR services for the reconstruction of 3D content or data based on ingested 2D content or data need to be in compliance with the **conditions of applicable exceptions and limitations** where such acts are not authorised in some other way.

Where the XR service is of such a nature that it comprises an automated analytical technique that generates information in the form of 3D content or data, it may be enabled by the TDM exceptions vis-à-vis underlying 2D content or data. Crucially, the **conditions of the TDM exceptions must be duly taken into account**. The availability of the TDM exceptions for such services will further depend on the nature and form of the ingestion stage. For instance, where an opt-out has been expressed under Article 4(3) of the CDSM Directive, that opt-out may still be infringed where the particular opted-out content or data is not obviously utilised (for instance, where the output relies on a significantly large number of inputs). This can be case whether or not the relevant content or data can be recognised in the output 3D content or data.

In that regard, the lawfulness of the ingestion stage is essential for the subsequent lawfulness of e.g., 3D reconstruction services. Furthermore, should the 3D reconstruction services have any form of intermediate stages that would necessitate a temporary copying of the underlying content or data, **the requirements of the temporary reproduction exception** must be complied with.

[105]Council of the European Union, 'Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World' <https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>.

7.4.3.1.3 Copyright: Licenses – Licensing of Copyright-Sensitive Acts in the Context of XR Services Implementation (WP4 Services) (AI-Generation license)

Ingested content, i.e. 2D content uploaded at the NMR storage, including content analysed via the implementation of WP3 Services, may be further used in the course of XR services provision (WP4 Services) for the purposes of new 3D assets generation, mainly involving algorithmic training[106]. URLs will be passed on to client XR services needing access to content.

In the course of implementation of XR services, the transient RAM copies as well as temporary storage of the ingested content in the respective server of the service that are expected to take place **are qualified as reproductions** of the ingested content[107], thus, in principle[108], subject to licensing.

Licensor should be the rightsholder(s) of copyright and/or related rights over the protected (analysed) content or a license holder of the reproduction right (in this case: sub-licensor). The (sub)licensee should be the XReco service provider.

7.4.3.2 Focus: New and Derivative Works in the Context of XR Services

In the XReco context, generation of 3D content is always based on pre-existing 2D content[109]. A key issue for the deployment of XR services, beyond the algorithm training of ingested content, is the nature of the relationship between underlying 2D content (ingested content) and resulting 3D content (New 3D asset). In copyright terms, this entails establishing whether the resulting 3D content in some way contains protectable expressions of the ingested 2D content (**Derivative Work**)[110].

At the same time, any XR service that generates 3D content using underlying 2D content may or may not create a copyright-protectable work in its own right. This depends on whether the subsistence criteria for works are fulfilled, such that a new work may be considered to have been created, for instance by the user of XR services, such that they are an author of the 3D content (New 3D Asset as **New Work)**. More specifically, copyright protects works of authorship where: (1) there exist an original subject matter that is the "author's own intellectual creation", reflecting "the personality of its author, as an expression of his free and creative choices"[111], and; (2) the work is classified as "the elements that are the expression of such creation"[112], meaning that there exists "a subject matter that is identifiable with sufficient precision and objectivity"[113]. 3D content in digital form, for instance as a file exported from the XR service, will practically always qualify as subject matter with sufficient precision and objectivity. This means that the key question of whether the 3D content is a new work is whether it **reflects the (human) author's personality as the expression of free and creative choices**. Importantly, originality is not achieved "[when] an expression is determined by technical or functional rules, such

[106] This is the word choice of the project partners (see Deliverable D4.1).
[107] Covered by Art. 2 InfoSoc Directive.
[108] Cf. Rec. 18 CDSMD: "Rightsholders should remain able to license the uses of their works or other subject matter falling outside the scope of the mandatory exception provided for in this Directive for text and data mining for the purposes of scientific research and of the existing exceptions and limitations provided for in Directive 2001/29/EC."
[109] I.e. it is not an *ex nihilo* endeavor, such as the creation of a 3D model "from scratch".
[110] Note however, that this issue is separate from the issue of underlying works contained in the 2D content, which is addressed in the context of data and content ingestion as well as from the issue of use of pre-existing content as training material for the development of XR Services AI models.
[111] *Painer*, para. 88, 89, and 94.
[112] *Infopaq I*, para. 37 and 39.
[113] *Levola*, para. 40; *Cofemel*, para. 32.

as when there is only one way to express an idea, or the expression is predetermined by a specific goal or constrained by narrow rules that leave no space for free and creative choices no originality can be present"[114].

EU copyright *acquis* does not contain rules of authorship or copyright ownership. However, it may be safely derived from the concept of originality that the concept of *authorship* and, consequently, the protectability of a work at EU law level is connected **with human *creation***[115]. Indeed, EU copyright law requires at least some human involvement in the process of creation for a work to be copyrightable[116]. Accordingly, EU law follows the traditional pattern of copyright protection and ownership that flows from human authorship[117]. As a consequence, **only human-authored AI-generated New 3D assets** may be copyrighted under EU law.

It should be noted, however, that the fulfilment of authorship requirement (and, thus, copyrightability) may be challenging in case of New 3D assets generated by XR services: authorship and, consequently, ownership could be attributed to **one or more human agents** involved in the creative algorithmic process (programmer/coder, data selector, trainer, etc.[118] and/or to an end user who interacts with the machine). At the same time, there may be the case that no authorial creative decisions originate from human agents. In this case such decision may rely (partially or completely) upon the machine itself[119], such as creative decisions that are unexplainable[120] and unpredictable/ unforeseeable (but not accidental or random) even for the programmers of the system[121] (**XR Services as AI Black boxes**). In case that no human authorship is established, the output will not be an original copyrighted work[122]. There is also still the case that human authorial decisions are connected with only some elements of the output. Then it should be assessed whether algorithmic participation in the creative process outshines the human involvement[123].

Nevertheless, the question emerges whether the generated 3D content is additionally/separately a "derivative work" – a work "that [is] based on pre-existing works" [124]. Under Art. 2 para. 3 Berne Convention, protection over derivative works is offered "without prejudice to the copyright in the original work". This would entail that the exclusive rights of the author or rightsholder of the underlying 2D content may be infringed by resulting 3D

[114] Thomas Margoni, 'The Harmonisation of EU Copyright Law: The Originality Standard' in Mark Perry (ed), *Global Governance of Intellectual Property in the 21st Century: Reflecting Policy Through Change* (Springer International Publishing 2016), 95; citing *Football Association Premier League*, para. 98; *Bezpecˇnostnı´ softwarova asociace*, para. 49; *Football Dataco*, para. 39.

[115] EP 2020, no 15: "[...] considers that works autonomously produced by artificial agents and robots might not be eligible for copyright protection, in order to observe the principle of originality, which is linked to a natural person, and since the concept of 'intellectual creation' addresses the author's personality."

[116] Martin Senftleben and Laurens Buijtelaar, 'Robot Creativity: An Incentive-Based Neighboring Rights Approach' (October 1, 2020). available at: <http://dx.doi.org/10.2139/ssrn.3707741>, 8.

[117] See among others, Christian Hartmann et al, 'Trends and Developments in Artificial Intelligence Challenges to the Intellectual Property Rights Framework. Final report,' Study Commissioned by the European Commission, prepared by The Joint Institute for Innovation Policy and University of Amsterdam [2020], available at: https://www.ivir.nl/publicaties/download/Trends_and_Developments_in_Artificial_Intelligence-1.pdf, 95: "Copyright ownership follows authorshiIt is a universal rule of copyright law that, by default, copyright vests in the person having created the work."; Enrico Bonadio and Luke McDonagh, 'Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity', (2020) 2 Intellectual Property Quarterly, 2.

[118] For this categorization of human agents in the up-stream algorithmic creative process, see Bonadio and McDonagh op.cit., 13.

[119] Annemarie Bridy, 'Coding Creativity: Copyright and the Artificially Intelligent Author' (2012) 5 Stanford Technology Law Review, 2.

[120] On the "explainability" question related with AI systems, see *Hartmann* et al., op.cit., 24.

[121] See Daniel Gervais, 'The Machine as Author' (2019) 105 Iowa Law Review 2053, 2070.

[122] EP 2020, no 15.

[123] Geiger and Iaia, 5.

[124] Thomas Margoni, 'The digitisation of cultural heritage: originality, derivative works and (non) original photographs' (IViR, 2015), available at: <https://www.ivir.nl/publicaties/download/1507.pdf>, 18; Cf. Section 1 (a) of Creative Commons License Attribution-NoDerivatives 4.0 International : "Adapted Material means material subject to Copyright and Similar Rights that is **derived from or based upon** the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor."

content. In the absence of an EU law definition, the term "derivative work" is used here without prejudice to specific provisions of national law of the EU Member States and may be subject to various modalities beyond the scope of this deliverable. The potential status of the resulting 3D content (New 3D asset) **may or may not therefore be either a new work and/or a derivative work.**

7.4.3.2.1 The "Dimension Shifting" Factor in the Creation of 3D Content: A Copyright Law Analysis

Creation of 3D content based on 2D content is a process which involves "**dimension shifting**", namely, the rendition of representations contained in the pre-existing 2D content in a visual format featuring an additional dimension, by means of XR services. In the passage from two to three dimensions, there is a need of modifications in 2D content, involving **addition of aspects, views, angles, perspectives**. 3D content may result from (a) a simple (i.e., slavish or imitative) incorporation (reproduction) (or juxtaposition of incorporations) of 2D content in the 3D content or (b) from a transformative rendition of 2D content. The first case refers to the category of "composite works" recognized by several national laws. Accordingly, the emerging question is whether dimension shifting results in an original and/or derivative New 3D Asset.

Dimension shifting & derivative work generation: Copyright protection over 2D content would extend to 3D content, and, therefore, dimension shifting would result in the creation of a derivative 3D work, insofar the 3D content contains elements from concrete exteriorized protectable expression of the 2D content[125]. This implies a first high-level check, i.e. the implementation of the idea/expression dichotomy (see Annex VI).

The recognizability criterion: In this context of derivative works and from an EU law perspective, in the absence of a harmonized adaptation right, the most relevant harmonized exclusive right is the right of reproduction. In that regard, a subsequent check for assessing the existence of a derivative works derives from the external limits of the right of reproduction under EU law[126], as defined by the CJEU. In particular, the CJEU, in *Pelham* ruled that the use of a sound sample from a phonogram in a modified form **unrecognisable** to the ear, in a new work does not constitute 'reproduction' within the meaning of the right of reproduction of phonogram producers[127]. In light of this approach, a similar delimitation of author's right of reproduction is conceivable[128], though without judicial confirmation. In the XReco context, this criterion would mean that the use of a pre-existing 2D content **in a modified form, unrecognizable to sight/the eyes**, in a new, 3D output, would not be qualified as reproduction and **would remain free**. Inversely, if recognisable elements of (one or more) pre-existing 2D content that enjoy copyright protection are incorporated in the new 3D content and recognizable to the sight, constitute a partial reproduction, subject to license[129]. In that regard, recognizability may then be correlated with the concept of "relevance" of contributing assets and its measurement/evaluation.[130]

The "distance" criterion: Another possible criterion for assessing the existence of a derivative work in 3D content may be drawn from national laws' traditions (e.g. the German doctrine of "freie Benutzung"), according to which the use of elements of pre-existing works in new works does not require prior authorization, insofar the new work reflects originality to an extent that **incorporated pre-existing elements "fade away"** and new work has a sufficient distance to, and is thus independent vis-à-vis, (a) pre-existing work(s). This approach requires a

[125] Cf. Judgment of 16 July 2009, *Infopaq International* (C-5/08, EU:C:2009:465), paragraph 39.
[126] Art. 2(a) InfoSoc Directive.
[127] A related right established by Article 2(c) of InfoSoc Directive; Case C-476/17 Pelham, para. 31 and 39.
[128] In that sense, K. Grisse and C. Kaiser, 'On the significance of (un)recognisability for the reproduction right in European copyright law' (2022) 44(2) EIPR 78, 81 and further references.
[129] Martin Senftleben, 'Flexibility Grave – Partial Reproduction Focus and Closed System Fetishism in CJEU, *Pelham*' (2020) 51 IIC 51, 751. This situation may also be referred as "traceability" of pre-existing (ingested) content within the New 3D assets, *cf.* EP 2020, Rec. D: "AI technologies may render the traceability of IPRs and their application to AI-generated output difficult, thus preventing human creators whose original work is used to power such technologies from being fairly remunerated."
[130] See DoA, 9.

comparison between the originality of used pre-existing elements and the originality of the expression of the new work. In the XReco context, the application of this criterion involves a comparison between the degree of originality of ingested content (which includes the originality of the represented theme) as well as the originality of the New 3D asset.

The above two criteria may inform each other: the less recognizable the pre-existing 2D works, the more distanced, and, thus, independent, the new work. In that regard, dimension shifting may be a key element in supporting such distance. However, even in the absence of derivative work, the 3D asset may still reproduce protected elements of the item which is depicted/represented in the (distanced or not recognizable) 2D pre-existing content (e.g., 3D content reproducing the appearance (form) of a protected architectural work but not the 2D photograph thereof).

7.4.3.2.2 Dimension shifting and originality

In the XReco context, it need to be checked whether the **addition of aspects, views, angles, perspectives** in 2D content, which are necessary for realizing the passage from two to three dimensions, constitutes an original (protectable) contribution. As in the case of digitalization of 2D content, it can be argued that the factor of "dimension shifting" is a key element in determining an original contribution, such that photographing 3D objects commonly implies more free and creative choices than photographing two-dimensional objects[131]. Inversely, when there is only one way to express an idea, or the expression is predetermined by a specific goal or constrained by narrow rules that leave no space for free and creative choices, such as e.g., in case of the identical (photorealistic) 3D reconstruction of external views of a monument, the originality standard would probably not be met.

Based on the above analysis, there are four possible separate statuses of the resulting 3D content, depending on the factors: (1) **applicable law** and its regulation of derivative works; (2) the **originality** of the 3D content. The status of each New 3D Asset is crucial as it affects (a) eventual rights clearance for their lawful commercialization and (b) their protection status.

7.4.3.2.3 Status No 1: New 3D Asset is New Work and Derivative Work

The 3D content may be simultaneously a new work and a derivative work. This is the case where the applicable law determines that the rights held in the relevant 2D content cover also 3D content that is based on this 2D content and where the author of the 3D content also enjoys protection due to his original contribution in creating the 3D content (cumulative creativity)[132]. 3D content would then be protected under copyright (a) over pre-existing work plus (b) copyright over derivative contribution (layered rights).

[131] Margoni https://www.ivir.nl/publicaties/download/1507.pdf, 51.

[132] Th. Chiou, Copyright law and algorithmic creativity: Monopolizing inspiration? σε Synodinou, T., Jougleux, P., Markou, C., Prastitou-Merdi, T. (Eds.), EU Internet Law in the Digital Single Market, (Springer 2021), 265 ff. and esno 11.

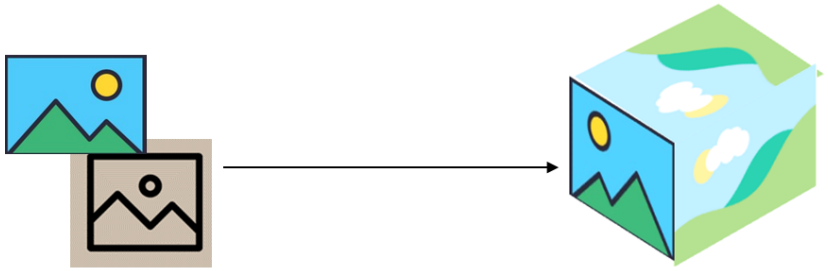


Figure 7-2 Recognisable 2D assets in a created 3D asset with creative choices.

An example of this may be 3D content that contains **recognizable protected elements** of protected pre-existing 2D content while, at the same time, the dimension shifting involves free and creative choices e.g., in the rendition of 2D content in 3D or in the arrangement of the incorporation of pre-existing content into the New 3D asset.

7.4.3.2.4 Status No 2: New 3D Asset is not a New Work but a Derivative Work

The 3D content may be solely a derivative work. This is the case where the applicable law determines that the rights held in the relevant 2D content cover also 3D content that is based on this 2D content, but where the creator of the 3D content does not enjoy protection, because his contribution fails to meet the threshold of originality. In some circumstances, the 3D content may be regarded as a copy of the 2D content. In that scenario, only rights over pre-existing content apply over the 3D content. This will be also the status for any derivative New 3D asset, where human authorship cannot be established.

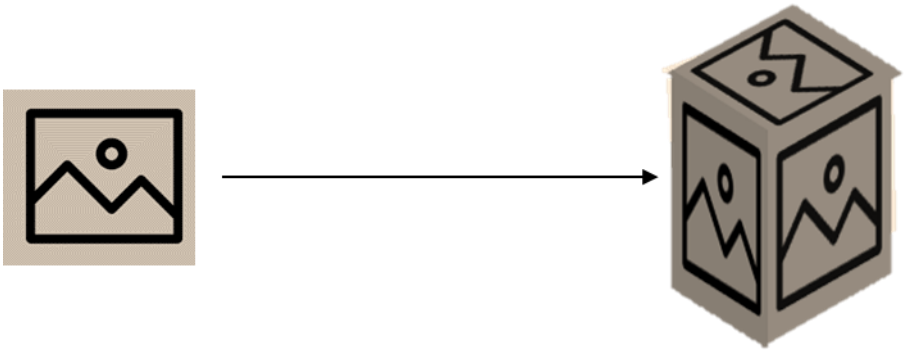


Figure 7-3 Recognisable 2D assets in a created 3D asset, created without creative choices.

An example of this may be 3D content that contains **recognizable protected elements** of protected pre-existing 2D content while, at the same time, the dimension shifting does not involve free and creative choices, for instance, where the creator of the 3D content has not made any adjustments to its arrangement or rendition.

7.4.3.2.5 Status No 3: New 3D Asset is New Work and not a Derivative Work

The 3D content may be solely a new work that meets the threshold of originality. This is the case where the applicable law determines that the rights held in the relevant 2D content do not extend to the 3D content that

is based on this 2D content, but the creator of the 3D content contributed originally to the production of the output. This will possibly be the case when the New 3D asset contains **unrecognisable fragments** of the pre-existing 2D content or that is sufficiently distanced from the pre-existing 2D content, while its expression reflects free and creative choices of the author. In that scenario, only 3D content rights apply.

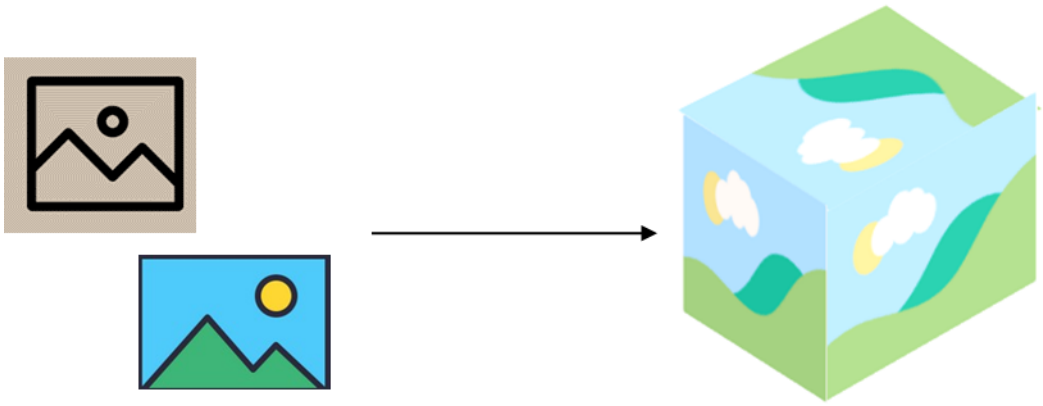


Figure 7-4 Created 3D asset without recognisable 2D source assets, but with creative choices.

An example of this may be 3D content that **does not contain recognizable protected elements** of protected pre-existing 2D content while, at the same time, the dimension shifting involves free and creative choices e.g., in the rendition of 2D elements in 3D or in the arrangement of the incorporation of pre-existing content into the New 3D asset.

7.4.3.2.6 Status No 4: New 3D Asset is neither New Work nor a Derivative Work

Last, there is a scenario where the 3D content is neither an original new work nor a Derivative Work. This will be the case when only **non-protected elements from pre-existing 2D works are expressed in the New 3D asset** while, at the same time, the contribution of the creator of the **3D content fails to meet the threshold of originality**. This status will apply to any non-derivative New 3D asset, where the human authorship cannot be established. In this scenario, 3D content would be totally unprotected by copyright law.

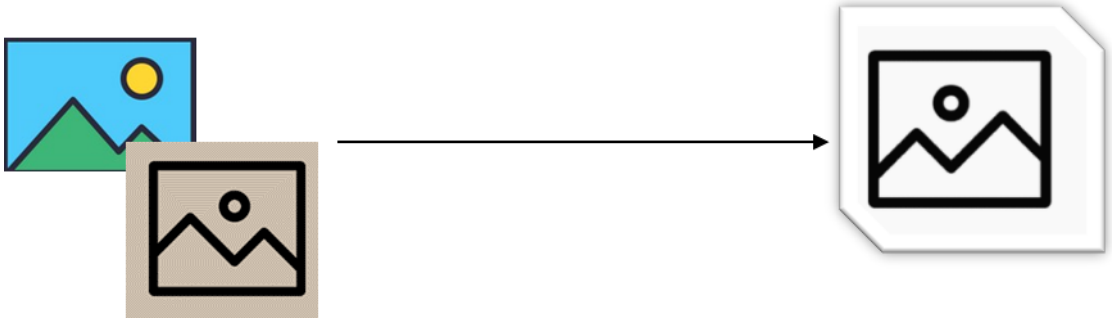


Figure 7-5 Created 3D asset without recognisable 2D source assets, created without creative choices.

An example of this may be 3D content that **does not contain recognizable protected elements** of protected pre-existing 2D content while, at the same time, the dimension shifting does not involve free and creative choices, for instance, where the creator of the 3D content has not made any adjustments to its arrangement or rendition.

7.4.3.2.7 Statuses of New 3D Assets

The matrix of possible combinations as outlined above, indicating the resulting outcome in terms of copyright protection of the 3D content in question is the following:

	New Work	No New Work
Derivative Work	Multiple Rights in 3D Content	No Rights in the 3D Content independent of subsisting 2D Content rights
No Derivative Work	Rights Held only in 3D Content	3D Content Unprotected

7.4.3.3 Analysis: Risks and Opportunities

Risks: At XR services stage, ingested content needs to be copied for technical reasons, mainly on a transient or temporary fashion. Still, reproduction right is triggered and the service provider that undertakes those acts (even if triggered by the user) needs to hold reproduction rights over ingested and analysed content, being noted that these reproductions are separate acts of reproduction from initial ingestion copy and that the reproduction made for ingestion purposes is expected to be made by the uploader (and not the service provider). Where specific XR Services rely on such intermediate stages within their functioning such that copies of the ingested 2D content or data are made, the requirements of the temporary reproduction exception must be complied with where such uses are not otherwise authorized.

There are chances that the New 3D asset is qualified as Derivative Work under applicable law, which in principle requires **rights clearance over pre-existing 2D content for that derivative use** (and beyond), given that the list of exceptions and limitations to exclusive rights set out at EU level, including reproduction right, does not include a general exception permitting the use of works of others for the purposes of creating a new work[133]. Sub-licensability of cleared rights on pre-existing 2D content is crucial for maintaining integrity of rights clearance at the Marketplace stage. Derivative creativity needs to be assessed on a **case by case basis**.

In addition, the assessment of originality of the New 3D asset remains additionally challenging, to the extent that the **attribution of human authorship**, if any, **is not straightforward**.

Opportunities: The conditions of exceptions and limitations to copyright are decisive for their availability in the context of the XR Services stages. This will also depend on aspects of the Ingestion stage. Especially, the TDM exceptions may be usable where certain XR Services, such as NeRF, qualify as automated analytic techniques.

All licensing needs at XR Services stage may be addressed by means of one license, which, however, needs to address the eventual need of further sublicensing. It may be expected that t**he input license (see above) over ingested 2D content anticipates the eventual processing at this stage and covers equally** all relevant licensing needs for the implementation of any WP4 services vis-à-vis ingested content (including the creation of derivative work). New rights over New 3D content are born at this stage.

All these licensing challenges are usually solved (at least, prima facie) when content is subject **to some types of open license** that allow the creation of derivative works (such as Creative Commons licenses without ND limitation). Measurement of relevance under could be linked with recognisability criterion and rights management at New 3D asset generation stage (prior to monetization).

[133] In that sense, see Opinion of Advocate General Szpunar in Case 476/17 *Pelham* [2019], para. 54.

Key Points

- Generation of New 3D assets by means of XR services should be lawful.[210]
- For the generation of New 3D assets by means of XR Services, the text and data mining exceptions may be available where the respective conditions are fulfilled.
- XR Services that involve intermediate stages for the generation of a New 3D assets, the temporary reproduction exception may be available where its conditions are fulfilled.
- Implementation of WP4 services typically involves one or more acts of reproduction that need to be previously licensed, unless if exceptions or limitations apply.
- These acts are expected to be performed by the service provider and not the user.
- New 3D assets may be either a new work or not and/or a derivative work or not. Applicable law offers criteria for legal assessment, which however needs to take place on a case-by-case basis, based on national laws.
- Whether New 3D assets enjoy copyright protection themselves depends on their originality.
- Open licensing possibly offers a promising licensing solution.

7.4.4 XR Marketplace

7.4.4.1 Key Legal Aspects

7.4.4.1.1 Copyright: Exceptions and Limitations

The operation of a marketplace for 3D content and data usable for XR application that are either reconstructed using the aforementioned XR services or pre-existing raises certain issues, specifically regarding reconstructed 3D content and data. In this regard, it ought to be underlined that where the 3D content or data created (New 3D Asset) is not a derivative work and solely a new work in light of the analysis in Section 7.4.3.2, **permission or authorisation for the downstream marketing of this content or data is not required**, assuming the person offering the content or data on the marketplace is the creator of that New 3D Asset. Where the New 3D Asset is not a new work, it may still be offered on the marketplace in certain ways, though it would not enjoy copyright protection.

Where the New 3D Asset is still a derivative work, its offering on the XR marketplace may be problematic. The XR content and data lifecycle might benefit from the **temporary reproduction exception** going from ingestion to the achievement of an XR service for 3D reconstruction. However, the final criterion, namely that "that act does **not have any independent economic significance**" would arguably be imperilled by offering the 3D reconstructed content or data on a marketplace. The attempt to monetize the New 3D Asset would manifest its economic significance. This could have implications for potential lightweight 3D reconstruction techniques that would hope to benefit from the temporary reproduction exception.

For cases where the New 3D Asset is still a derivative work and where the reconstructed 3D content and data hopes to benefit from the TDM exceptions, the placing on a marketplace of such 3D content and data may nevertheless be held unlawful where it is found to go against the **three-step test**[134], meaning that the exceptions "shall only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightsholder"[135]. Whether

[134]For the TDM exceptions, Article 7(2) CDSMD clarifies that "Article 5(5) of Directive 2001/29/EC shall apply to the exceptions and limitations provided for under this Title".
[135]Art. 5(5) InfoSoc Directive.

or not (1) the 3D reconstruction of underlying 2D content or data is sufficiently specific a case; (2) it forms part of the normal exploitation of the work or other subject matter, and; (3) it unreasonably prejudices the legitimate interests of the rightsholder, is not certain. This arguably depends on a **case-by-case analysis**, specifically on the chain of events in the **XR production lifecycle**, the **acts performed on underlying 2D content or data**, the **legal bases** (exceptions, limitations, and/or licenses) relied upon for 3D reconstruction, and generally, **unsettled questions of EU law** or matters specific to national law.

7.4.4.1.2 Copyright: Licenses

7.4.4.1.2.1 Marketplace license

At the moment, the exact technical features of the marketplace are not yet precisely established. However, any marketing (sharing) of New 3D assets via an online marketplace requires **storage of the file that embodies the New 3D asset**, in order to make possible the display (making available) of that asset at the marketplace website. The storage of the New 3D asset would be qualified as reproduction of the embodied 3D asset (including reproduction of the pre-existing work(s) expressed in the New 3D asset)[136], and would be, in principle, subject to licensing.

Moreover, the eventual **public display** of marketed New 3D asset at XReco Marketplace website, even if the latter is to be seen as a P2P network, will constitute an act of communication to the public[137], and, equally, (in principle) subject to licensing.

Licensing needs vary, depending on the status of 3D content described above and, therefore, the rights that need to be cleared.

7.4.4.1.2.2 End-user license / Reuse license

The marketed New 3D asset, if qualified as a copyrighted work, needs to be accompanied by an **end-user license** that will define the terms of its use and/or exploitation by the interested end-users. The definition of the terms of this license may be freely negotiated by the parties or may be predefined (e.g., licensing schemes proposed by the Marketplace operator, open licenses such as Creative Common licenses etc.), subject to the principle of freedom of contract (see Annex VII), without prejudice to:

a) Mandatory EU copyright contract law (see Annex VII)

b) Respect of rights over pre-existing works, in case New 3D Asset is a Derivative Work (and subsequently pre-existing work(s) contained in the New 3D asset are further used, displayed, exploited etc.) (**reuse license**).

Under both licensing instances (Marketplace display and End-user license), Licensor for relevant rights should be the rightsholder(s) of copyrights over the New 3D asset and/or the rightsholder(s) of copyrights over pre-existing works(s) expressed in the New 3D asset (in case of Derivative Work) or a license holder of the relevant rights. In case of Derivative Work, the licensor will act as (sub)licensor equally for the rights over pre-existing works that are expressed in the New 3D asset.

In case of display license, the licensee may be the operator of XReco marketplace website, depending on its legal status (**intermediary**)[138], while for the act of uploading as such the user (uploader) also need to be licensee (in

[136] Covered by Art. 2 InfoSoc Directive.
[137] Covered by Art. 3 InfoSoc Directive.
[138] Cf. Art. 17(1) CDSMD.

the reuse license context)[139]. In case of end-user license attached to the New (marketed) 3D asset, the licensee shall be the **end user** who acquires the New 3D asset. Further use/exploitation of the New 3D asset (and all possible rights involved) should equally be covered by the end-user license**. End users** in the XR marketplace may benefit from exceptions and limitations to copyright in regard to New 3D assets that are new works. In light of their statutory nature, an end-user license can clarify how and in what form end users can enjoy these exceptions and limitations.

7.4.4.1.3 Own use by the XReco User

The stage of XR Marketplace may be skipped in case that the user who prompted the New 3D Asset generation at XR Services stage intends to proceed with use or exploitation for own purposes. Still, such use would in principle require prior clearance of third-party rights over the New 3D asset (in case of original work featuring authorial contributions of human agents other than the user) and/or rights over pre-existing works that are possibly expressed in the New 3D asset (in case of Derivative Works). In that case, the owners of third-party rights will be licensors and the user will be the licensee. No licensing needs arise in case that user triggered the generation of a New 3D asset **based on own ingested content**.

7.4.4.2 Analysis: Risks and Opportunities

Risks: The exploitation of New 3D Assets that are derivative works is generally highly problematic when relying on exceptions and limitations, as such a use of underlying 2D content is unlikely to satisfy the three-step test.

Sharing of New 3D Assets in the XR Marketplace and further use by end-users or the user that triggered its generation by means of XR Services involves copyright-sensitive acts that possibly affect several layers of rights, depending on the copyright status of New 3D Asset.

Where the New 3D Asset is a Derivative Work, lawful use/exploitation implies as a minimum the compatibility (sufficient coverage) of the scope of cleared rights over pre-existing 2D content on the one hand, and the terms of use attached to the New 3D content on the other[140]. This means that the terms of use regarding Derivative New 3D content may not exceed the scope of licensed use of ingested 2D content.

Also, New 3D Asset needs to be covered by the same open license, if pre-existing content is subject to an open license featuring a “share alike” clause, such as Creative Commons licenses with an SA component.

The nature and role of XReco as a service provider or intermediary at the Marketplace stage may affect rules applicable over licensing agreements that cover XReco service provision.

Opportunities: Where the New 3D Asset is not a derivative work and its offer or is the same person as the owner of copyrights in the new work, that person is permitted to offer the New 3D Asset on the XR marketplace. Where the New 3D Asset is not a derivative work but also not a new work, the content or data may still be offered on the XR marketplace in some way.

Given that freedom of contract remains the default rule, the exact content of end-user licenses these agreements could be determined by means of calibration of the main licensing building blocks described in Annex VIII, with

[139] Cf. Art. 17(2) CDMSD; Cf. also R. Mittal, ‘Mechanisms to Make End-Users of Copyrighted Works Pay Through Levy and DRM’ in Kung-Chung Liu and Reto M. Hilty, (ed.), Remuneration of copyright owners: regulatory challenges of new business models (Springer 2017), 122.

[140] Cf. EP 2020, no 15.

due respect to the mandatory applicable law provisions and the licensing needs that characterize all stages of XReco workflow.

Key Points

- New 3D Assets that are not derivative works may be economically exploited and offered on the XR marketplace by their authors.
- New 3D Assets that are not derivative works and are not new works may still be offered on the XR marketplace in some way.
- The economic exploitation of New 3D Assets that are derivative works may be hazardous for their creators, and unlikely to benefit from the exceptions and limitations to copyright in light of the three-step test.
- Placement and sharing of protected New 3D Assets at the XR Marketplace entails acts of reproduction and communication to the public. Licensing solutions depend on the status of the intermediary/platform.
- All rights over marketed New 3D Assets, which may include rights over pre-existing content, in case of Derivative Works, need to be sufficiently cleared prior to placement/sharing.
- Any marketed New 3D Asset needs to be accompanied by an end user license, which should cover also rights over pre-existing content of Derivative 3D Assets.

7.5 Overall Insights

- Data and Content Ingestion
  - Ingested data and content may contain personal data. In order to avoid issues concerning data protection and privacy, it is best not to access or use data and content containing living real persons.
  - Ingestion of protected content typically comprises one or more acts of reproduction (including for purposes of NMR).
  - A differentiation in licensing needs exists between Metasearch ingestion and user upload ingestion.
  - Temporary reproductions may be permitted at the ingestion stage without a license, provided that the legal requirements for such reproductions are met.
  - Reproductions within the context of text and data mining may be permitted at the ingestion stage, provided that access to the content or data is lawful, and the specific further requirements are met.
  - All necessary licenses need to be granted prior to ingestion/display and must also cover all protectable elements into the licensed content.
  - Open content constitutes a promising alternative to rights clearance via individual licensing of (retrieved and) ingested content.
- Search and Retrieval
  - Search and retrieval of purely factual elements that are not original expressions do not trigger any copyright concerns.
  - The implementation of certain WP3 services may involve one or more acts of reproduction that need to be previously licensed, unless exceptions or limitations apply.
  - Where factual elements for the purposes of search and retrieval are extracted from protected content or data, the temporary reproduction or text and data mining exceptions may be available, subject to the fulfilment of their respective conditions.
  - These acts are expected to be performed by the XReco service provider and not the user.
  - Open content constitutes a promising alternative to rights clearance via individual licensing of analysed content.

- XR Services
  - Generation of New 3D assets by means of XR services should be lawful.
  - For the generation of New 3D assets by means of XR Services, the text and data mining exceptions may be available where the respective conditions are fulfilled.
  - XR Services that involve intermediate stages for the generation of a New 3D assets, the temporary reproduction exception may be available where its conditions are fulfilled.
  - Implementation of WP4 services typically involves one or more acts of reproduction that need to be previously licensed, unless exceptions or limitations apply.
  - These acts are expected to be performed by the service provider and not the user.
  - New 3D assets may be either a new work or not and/or a derivative work or not. Applicable law offers criteria for legal assessment, which however needs to take place on a case-by-case basis, based on national laws.
  - Whether New 3D assets enjoy copyright protection themselves depends on their originality.
  - Open licensing possibly offers a promising licensing solution.

- XR Marketplace
  - New 3D Assets that are not derivative works may be economically exploited and offered on the XR marketplace by their authors.
  - New 3D Assets that are not derivative works and are not new works may still be offered on the XR marketplace in some way.
  - The economic exploitation of New 3D Assets that are derivative works is hazardous for their creators, and unlikely to benefit from the exceptions and limitations to copyright in light of the three-step test.
  - Placement and sharing of protected New 3D Assets at the XR Marketplace entails acts of reproduction and communication to the public. Licensing solutions depend on the status of the intermediary/platform.
  - All rights over marketed New 3D Assets, which may include rights over pre-existing content, in case of Derivative Works, need to be sufficiently cleared prior to placement/sharing.
  - Any marketed New 3D Asset needs to be accompanied by an end user license, which should cover also rights over pre-existing content of Derivative 3D Assets.
- Overall
  - Given the nature and restrictive interpretation of exceptions and limitations and, especially, the broad definition of reproduction right, licensing may be seen as the **default solution** for rights management, whereas rights management **based on exceptions and limitations** is a **viable alternative in select cases**.

7.5.1 Further steps

Further aspects of the above analysis will be in need of further development, particularly with a view to the work regarding Deliverable D3.2. In that regard, future work will seek to address the following:

- The development of the XReco workflow, including the technical design and business model choices regarding: (1) content and data ingestion; (2) search and retrieval; (3) XR services, and; (4) the XR marketplace. This will seek to address concrete envisioned solutions directly.
- Aspects of exceptions and limitations to copyright and related rights that can be further developed and researched in light of the above developments.
- Aspects of licenses that can be further developed and researched in light of the above developments.
- The role of the incoming EU AI Act on aspects of the XReco workflow.
- A more comprehensive study of the focus on XR Services regarding the distinctions between Derivative Works and New Works in the context of 3D reconstruction techniques, including where aspects of EU Member State national law can serve informative functions. This will address more deeply the criteria of recognisability and of sufficient distance.
- Further guidance for stakeholders in the XReco data sharing platform.

7.6 First Version of Licensing Components

Given the above discussion, irrespective of the status of harmonization of the legal framework for licensing contracts, the implementation of licensing mechanisms/agreements in the XReco context will be still necessary

to the extent that exclusive copyrights or related rights are involved (and **insofar no exception or limitation are lawfully applicable**).

A summary of XReco licensing components, related with both pre-existing content and New 3D content, which may be part of the Rights management component may be the following:

XReco licensing components (v.1)

Necessary for ingestion:

Upload license (for pre-existing content)

Optional (if relevant services are used):

Display license (metasearch display) (for pre-existing content)

Description services license (for pre-existing content)

Necessary for content generation:

AI-generation license (for pre-existing content)

Necessary if content shared at the Marketplace:

Display license (marketplace display/sharing) (for copyrighted New 3D asset)

End-user license (for copyrighted New 3D asset)

Necessary if shared content is derivative creation

Reuse license (for pre-existing content)

Necessary for own use or exploitation of New 3D asset by XReco user

License for third-party rights (if new asset is co-authored or derivative)

There is an opportunity for XReco to set standards and establish fair and clear data retrieval, equitable remuneration of rightsholders, simplified licensing procedures and compliance with copyright regulations and ethical principles. Focus could be made on the technological identification of traces of expression of pre-existing works within new 3D assets as well as on defining metrics of significance of contribution, allowing machine-to-machine management and monetization of rights under the light of EU acquis.

7.6.1 Assessment of the Monetization Component from a Copyright Licensing Perspective

Licenses are typically granted by a licensor (rightsholder) to a licensee (user) against **remuneration**. Accordingly, in a licensing context, monetization stems from the granting of license and a license function as a means of monetization for the licensor (rightsholder)[141], **unless if parties (and especially, the licensors) opted for a free license** (such as Creative commons license)[142].

Accordingly, in the XReco context, monetization of licensed rights/licensed content to the benefit of licensor/rightsholder may be connected **with any, some or all licensing components** and may be refer to monetization of pre-existing (licensed) content and/or New 3D assets. In the first case of the remuneration functions as **revenue sharing entitlement** for pre-existing content creators (remuneration program). In the latter, the remuneration functions as revenue channel for XReco users. Of course, not all transactions reflect monetary value of equal importance (e.g., 'small licenses').

7.6.1.1 Monetizing pre-existing content

Monetization stemming from rights management over pre-existing content may in principle be connected with the below licensing components:

a) individual use of pre-existing content (upstream level):

- **Upload license[143]**
- **Metasearch display license**
- **Description services license**
- **AI-generation license**

b) reuse of pre-existing content (derivative creation) (downstream level):

- **Marketplace and reuse license (display 4and end-user license):** In case that the New 3D asset is a derivative work (pre-existing work(s) are expressed in the New 3D asset) or a derivative creation (pre-existing work(s) are expressed in the New 3D asset but the New 3D asset is not copyrighted as such) remuneration in favour of the rightsholders of pre-existing works could be agreed as exchange for licensing the reproduction, display (at the marketplace) and further use/exploitation of pre-existing works as parts of the New 3D asset.

7.6.1.2 Monetizing New 3D assets

Monetization stemming from rights management over New 3D assets may be connected with the below licensing components:

[141] It is in that regards that XReco rights management and monetization components are interconnected.
[142] In case of free license, monetization is in principle excluded.
[143] It should be noted that monetization would in theory be possible but probably would not be meaningful in case of ingestion (and further processing) by means of **user upload**, to the extent that the user is assumed to be the holder of rights over ingested content and therefore, being the licensor in the
se transactions, he would be remunerated for mere use of the services.

End-user license: In case that the New 3D asset is copyrighted, monetization may derive as exchange for the end-user license attached to this asset, upon listing (sharing) of this asset at the marketplace[144].

7.6.1.3 Fixation of remuneration

The fixation of remuneration remains largely subject to **contractual freedom**, without prejudice of applicable mandatory (copyright) contract rules (principle of adequate and proportionate remuneration for the exploitation of works and other protective licensing rules) which, however, apply in case that the licensor is the initial rightsholder (author or performer) of rights over pre-existing content or New 3D content (not in B2B or B2C context). For example, this will be the case of reuse and end-user licenses for the exploitation of New 3D asset consisting in derivative work. They would apply in case licensor is the initial rightsholder (author or performer) of rights over the ingested content.

Notwithstanding the limited scope of protective rules regarding the definition of appropriate and proportionate remuneration (in form of royalties-percentage or lump sum), some more general insights may be deducted, re: the factors that need to be taken into account for the definition of the remuneration in case of licenses granted for the exploitation of a work, such as the value of the licensed rights and the author's or performer's contribution to the overall work or other subject. This is relevant especially in the context of monetization of pre-existing works by means of reuse and end-user licenses for the exploitation of New 3D asset consisting in derivative work or co-authored work.

In that regard, according to the EU legislator (see rec. 73 and 77 Dir. 2019/790) the **significance** of author's or performer's **contribution** is an aspect that is considered, among others, as decisive, in order to determine an appropriate and proportionate remuneration for a license for exploitation. Accordingly, the factors for assessing the significance need to be defined in connection with the particular aspects of AI generated context of creation. In that regard, the "relevance algorithms" should be based on such assessment. The choice of relevance algorithm in the context of Monetization manager microservice development, where applicable, should be put under scrutiny vis-à-vis the contours of the concept of "significance of contribution" used under Dir. 2019/790 as a factor of fixating the remuneration for exploitation.

Last, remuneration of rightsholders in the context of licensing need not always be monetary but may also take the form of **non-monetary consideration**, such as the submission of data (such as data related with the use of content). Besides it may be argued that non-monetary consideration is also acceptable in case of primary licenses[145]. The assessment of significance of contribution would be used as a factor for fixing the remuneration of multiple rightsholders involved also in this context.

[144] Monetization would in theory be possible but probably would not be meaningful for the case of licensing the reproduction/display of New 3D asset at the XReco Marketplace, insofar the user that prompted the generation of the asset is owner of rights over this marketed asset.

[145] Theodoros Chiou, Non-monetary author's remuneration under EU Copyright Law: the case of data, REDA Conference, [conference presentation] 5 May 2023, Nicosia.

8 Rights management and licensing tools (FINC)

8.1 Overview

Rights management is a crucial aspect to preserve the interests of content creators and asset owners. Since technology offers unprecedented access to digital content, it is crucial to implement specific processes to identify, evaluate and exploit Intellectual Property (IP) strategies. Through mechanisms such as Intellectual Property Rights (IPR) management, rights management strikes a balance between promoting innovation, encouraging creativity, and safeguarding the economic value of IP. To manage the licences of the assets that will be uploaded into XReco, we need a strong licensing mechanism that can be created and validated automatically. In order to do this, we will use, as anticipated, Smart Legal Contracts (SLC).

A Smart Legal Contract is defined as: "a human-readable and machine-readable agreement that is digital, consisting of natural language and computable components. The human-readable nature of the document ensures that signatories, lawyers, contracting parties and others can understand the contract. The machine-readable nature of the document enables it to be interpreted and executed by computers, making the document 'smart'"[146]. Therefore, the SLCs links and automates the legal and physical aspects of contracting by translating a contract language into executable code. This code communicates with external sources, allowing self-execution based on input data. In this way, we can dynamically create and validate asset licences without manual operations.

But a SLC itself isn't enough to be legally valid as it must be "signed" and "stored" by third parties (e.g., notaries, certification authorities, etc.). To do so, we will use Blockchain Smart Contracts (SC) to notarize SLCs at the end of the SLC creation process. Smart Contracts are defined as "automated software agents hosted on blockchains that are capable of autonomously executing transactions on the triggering of certain conditions"[147]. Therefore, thanks to the infinite possibilities of SCs, it will be possible to notarize the data of an SLC onto the Blockchain to benefit from the advantages offered by the technology itself, such as immutability, visibility and transparency.

The Blockchain Network that will be used in the XReco project will be self-hosted by XReco partners. This choice is useful to avoid high transaction costs that are subjected to market trend. To do so, we plan to use Hyperledger Besu, an open-source Ethereum client, that allows us to create public or private blockchain (see Section 8.4).

Furthermore, in the early stages of the project, the blockchain will be permissioned to avoid the possibility by third parties to perform transactions not related to the project.

With respect to the SLCs we will, instead, refer to the Accord Project implementation[148].

8.2 Rights Management Architecture

In the architecture of XReco, there is a component dedicated to rights and licence management. This component contains various microservices such as Rights Management (RM), Monetisation Manager (MM), SLC Engine and Blockchain Service Provider (BCSP).

[146] Roche, Niall, et al. *"Ergo-a programming language for Smart Legal Contracts."* https://doi.org/10.48550/arXiv.2112.07064 (2021)
[147] Jake Goldenfein and Andrea Leiter. "Legal Engineering on the Blockchain: "Smart Contracts" as Legal Conduct" (2018)
[148] https://accordproject.org/

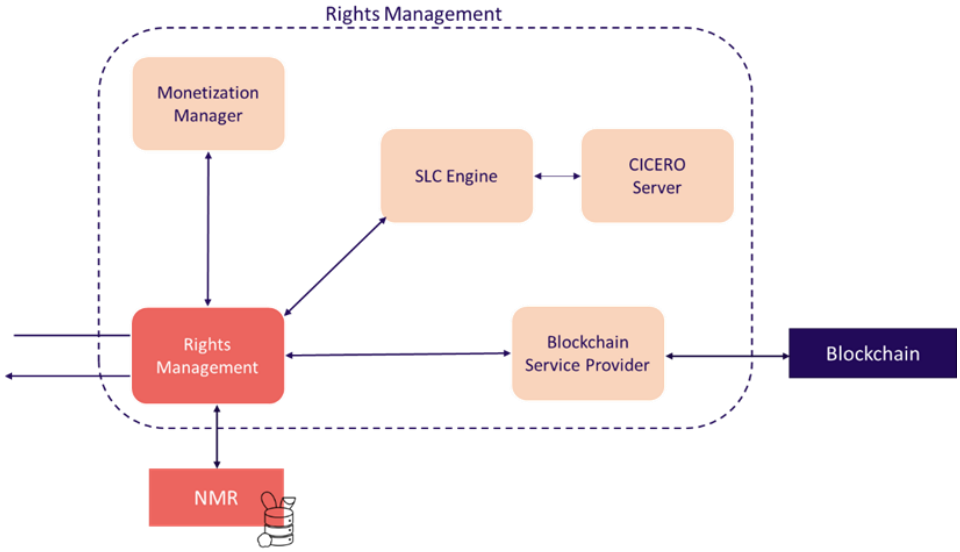


Figure 8-1 Rights Management Architecture.

More specifically, RM is the service that is exposed to the other components of XReco and is responsible for intercepting requests and managing the other microservices, and specifically:

- SLC Engine: that handles all aspects of SLC, such as creation, validation and execution.
- Blockchain Service Provider: that handles the communication with the Blockchain, such as deploying and interacting with SCs.
- Monetization Manager: that calculates the relevance of the assets used to create the new asset that is being registered.

During the entire process, RM makes multiple calls to the NMR Service to query and save data to the XReco database (Figure 8-2).

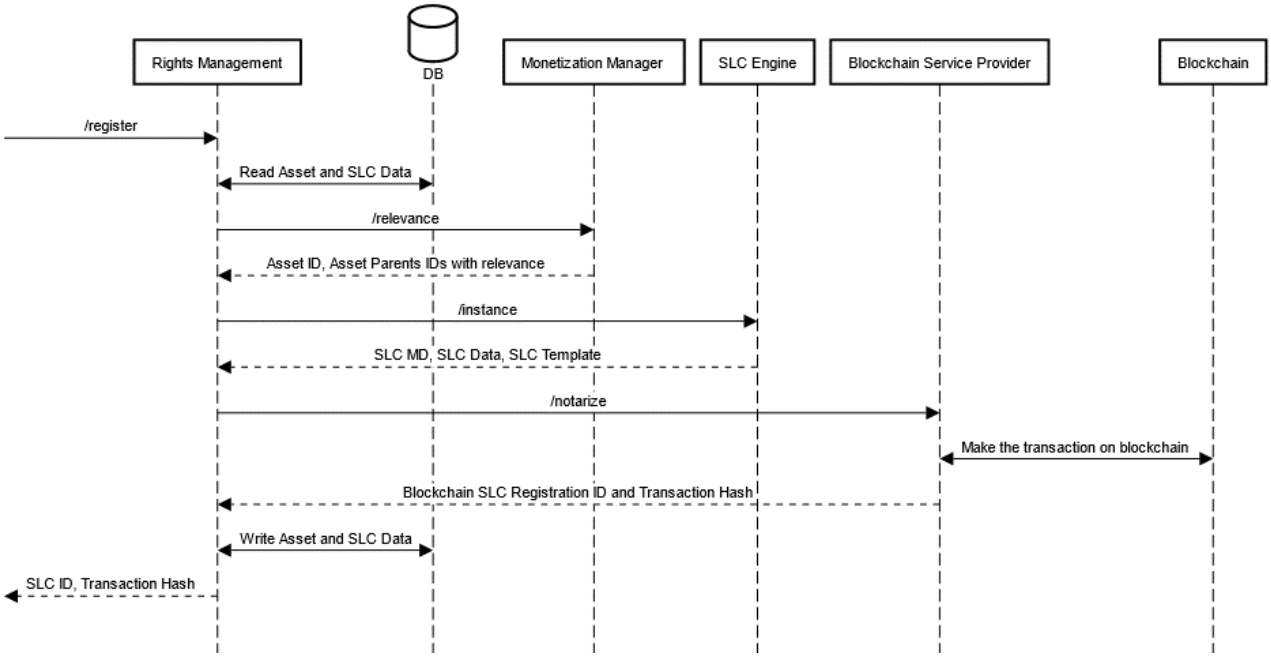


Figure 8-2 Rights Management and Monetization Registration Workflow.

To give an example of how the registration process takes place, Figure 8-2 shows a sequence diagram of a right license registration. First, the RM service retrieves the data concerning the asset from the DB. Subsequently, it calls the MM to calculate the relevance of the asset. Then, it calls the SLC Engine that returns the text and metadata of the SLC according to the licence requested. Finally, it calls the BCSP which notarises the metadata and hashes of the asset and the SLC to the Blockchain.

These services will be developed starting from their Application Programming Interfaces (APIs), i.e., the interfaces that allow software to interact and communicate with each other. The development process will therefore follow the API-First approach, which, as already mentioned in D2.1, involves the development process starts with the definition of the API, followed by the implementation of the methods.

To define the APIs, we will use the OpenAPI standard (OAS), which is a specification language for HTTP APIs that defines the structure and syntax independently from the programming language. Therefore, each service will have its own API specification defined in a YAML or JSON file. Thanks to this definition, it will then be possible to generate the methods (server and client) and web documentation. Such generators are available for numerous programming languages and frameworks, such as Android, Go, Python and Spring Boot. The OpenAPI specification of the Rights Management components are available in the project's Git.

For the development of all services (Rights Management, Monetisation Manager, Blockchain Service Provider and SLC Engine) we will use Spring Boot, which is a framework for building Java-based microservices and standalone applications with minimal configuration. Once development is complete, a Docker image will be created for every service, which can then be run on any machine, facilitating the testing and deployment phases.

In the following sections we provide specific details on the different components of the Rights Management architecture in XReco.

8.3 SLC Backend

As said before, SLCs represent a new paradigm within the realm of legal agreements, joining the traditional foundations of contractual law with the dynamism of computational technologies. Defined as machine-readable contracts, they encapsulate legal terms and conditions in a format suitable for automated execution. This evolution from paper-based contracts to programmable entities introduces a new dimension of efficiency, transparency, and enforceability into the legal landscape. SLCs, thus, aim to streamline and automate various aspects of contract management and execution, potentially reducing the need for intermediaries and enhancing the efficiency and transparency of legal processes.

8.3.1 Accord Project

To create SLCs we will use the Accord Project ecosystem. Accord Project is an open-source initiative, by the Linux foundation, that focuses on the development and implementation of smart legal contracts. It provides a framework and set of tools to create, manage, and execute legally binding contracts using blockchain and distributed ledger technologies. The project aims to standardize the way legal agreements are represented and executed, making them more accessible and interoperable across different platforms. The implementation of accord project templates is called Cicero. It includes three components (Figure 8-3):

- TemplateMark (TEXT): is the natural language of the clause or contract. It can include markup to indicate variables for that template.

- Concerto (MODEL): Accord Project templates associate a model to the natural language text. The model acts as a bridge between the text and logic; it gives the users an overview of the components, as well as the types of different components.
- Ergo (LOGIC): is the programming language which is used to express contractual logic in templates.

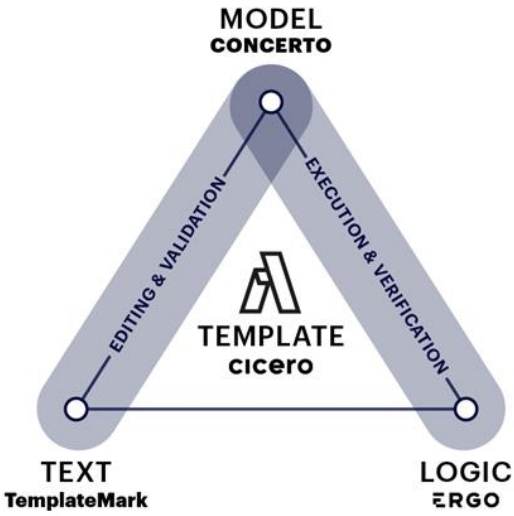


Figure 8-3 Accord Project's components.

This enables users to generate, verify, and execute software templates that encompass all three facets outlined in. As an example, the template creation starts by writing the text of the SLC:

```
1 Copyright License Agreement
2
3 This COPYRIGHT LICENSE AGREEMENT (the "Agreement"), dated as of {{effectiveDate}} (the
  "Effective Date"), is made by and between {{licensee}} ("Licensee"), a {{licenseeState}}
  {{licenseeEntityType}} with offices located at {{licenseeAddress}}, and {{licensor}}
  ("Licensor"), a {{licensorState}} {{licensorEntityType}} with offices located at
  {{licensorAddress}}.
4
5 [...]
6
7 {{#clause paymentClause}}
8 Payment. As consideration in full for the rights granted herein, Licensee shall pay
  Licensor a one-time fee in the amount of {{amountText}} ({{amount}}) upon execution of
  this Agreement, payable as follows: {{paymentProcedure}}.
9 {{/clause}}
```

Figure 8-4 Example of SLC Template.

Figure 6 shows the contract text with variables and clauses. Variables are identified between **{{** and **}}** and allows template to be used in different agreements by replacing them with different values (Figure 8-5).

```
Copyright License Agreement

This COPYRIGHT LICENSE AGREEMENT (the "Agreement"), dated as of 12/30/2017 (the "Effective Date"), is made by and between "Me" ("Licensee"), a "NY" "Company" with offices located at "1 Broadway", and "Myself" ("Licensor"), a "NY" "Company" with offices located at "2 Broadway".

[...]

{{#clause paymentClause}}
Payment. As consideration in full for the rights granted herein, Licensee shall pay Licensor a one-time fee in the amount of "one hundred US Dollars" (100.0 USD) upon execution of this Agreement, payable as follows: "bank transfer".
{{/clause}}
```

Figure 8-5 Example of full SLC Text.

Clauses, instead, are identified by the starting tag {{#clause clauseName}} and the ending tag {{/clause}} and allow users to customize the content of the SLC according to their need by selecting specifically pre-defined building blocks. For example, in the template in Figure 8-4 there is a paymentClause that can be included in the contract if a payment is foreseen between the licensee and the licensor.

To enable the text to be understood by the machine, Accord Project created an intermediate level called Model that is used to categorize text variables (number, monetary value, date, business or organisation, etc.): Figure 8-7

```
/* The template model  */
asset CopyrightLicenseContract extends Contract {
  /* the effective date */
  o DateTime effectiveDate

  /* licensee */
  --> Party licensee
  o String licenseeState
  o String licenseeEntityType
  o String licenseeAddress

  /* licensor */
  --> Party licensor
  o String licensorState
  o String licensorEntityType
  o String licensorAddress
```

Figure 8-6 Example of Accord Project's Model.

In Figure 8-6 the computer knows, for example, that the *licensee* variable (“Me” in Figure 8-5) and the *licensor* variable (“Myself” in Figure 8-5) are both *Party* types. The combination of Model and Template already make templates machine-readable. To complete the creation of the SLC template and make it also machine-executable, it’s necessary to write the Logic:

```
define function monthlyPaymentFormula(loanAmount: Double, rate: Double, loanDuration: Integer) : Double {
  let term = longToDouble(loanDuration * 12);       // Term in months
  if (rate = 0.0) then return (loanAmount / term)   // If the rate is 0
  else
    let monthlyRate = (rate / 12.0) / 100.0;        // Rate in months
    let monthlyPayment =                            // Payment calculation
      (monthlyRate * loanAmount)
      / (1.0 - ((1.0 + monthlyRate) ^ (-term)));
    return roundn(monthlyPayment, 0)                // Rounding
}
```

Figure 8-7 Example of Accord Project's Logic.

Each logic function has a name (e.g., monthlyPaymentFormula, Figure 8-7), a signature indicating the parameters with their types (e.g., loanAmount:Double), and a body which performs the appropriate computation based on the parameters[149]. In this example the function *monthlyPaymentFormula* in Figure 8-7 calculates the monthly rate value based on the other data in the text, i.e., the amount of the loan (loanAmount), the interest rate (rate) and the duration of the loan (loanDuration).

To write, create, edit and test SLC templates, Accord Project Template Studio comes to help. It is a web-based tool (Figure 8-8) that gives users an intuitive interface to design and visually refine templates.

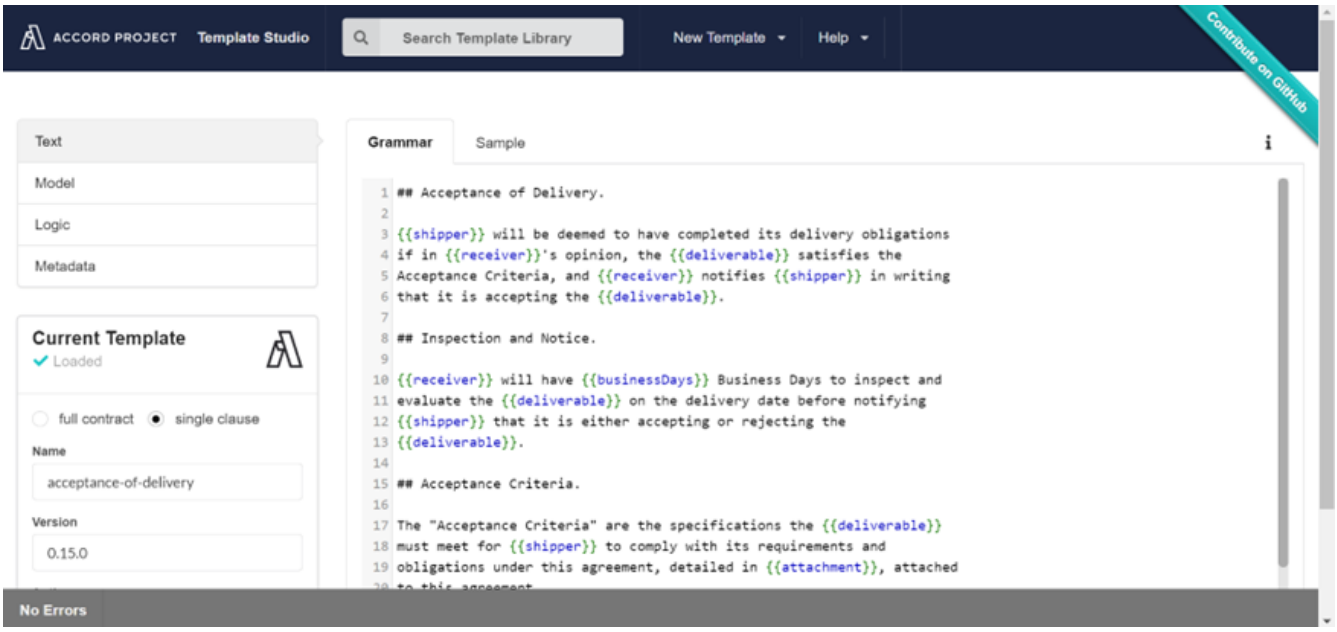


Figure 8-8 Accord Project Template Studio.

To automate the process of creating and verifying the SLC, Accord Project has developed different versions of Cicero to suit various needs:

- Cicero CLI: for parsing, executing, and creating archives via command line interface.
- Cicero Core: core classes to manage the grammar, models, and logic of Accord Project legal templates.
- Cicero Engine: a Node.js VM based execution engine for Accord Project legal templates.
- Cicero Server: that exposes Cicero Engine as a RESTful service.

[149] https://docs.accordproject.org/docs/accordproject-template.html

8.3.2 SLCs in XReco

SLCs will be instantiated and used to define the licensing of assets (with reference to the owners/licensee) uploaded inside the XReco platform. The XReco SLC Backend is the component responsible for the creation and validation of the SLC Text. To perform these operations and expose simplified APIs for Rights Management, SLC Engine uses a Cicero Server instance that contains all the SLC Templates supported by XReco.

Currently, the SLC Templates that are available for registration in XReco are:

- Creative Commons:
  - CC BY: *"This license enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use."*
  - CC BY ShareAlike (CC BY-SA): *"This license enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified material under identical terms."*
  - CC BY NonCommercial (CC BY-NC): *"This license enables reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator."*
  - CC BY NonCommercial ShareAlike (CC BY-NC-SA): *"This license enables reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator. If you remix, adapt, or build upon the material, you must license the modified material under identical terms."*
  - CC BY NonDerivs (CC BY-ND): *"This license enables reusers to copy and distribute the material in any medium or format in unadapted form only, and only so long as attribution is given to the creator. The license allows for commercial use."*
  - CC BY NonCommercial NonDerivs (CC BY-NC-ND): *"This license enables reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator."*
  - CC Zero (CC 0): *"is a public dedication tool, which enables creators to give up their copyright and put their works into the worldwide public domain. CC0 enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, with no conditions."*[150]
- Montreal Data License: offers a modular approach to data licensing in AI and ML. It provides various rights such as access to data, distribute the data, create a representation of data etc.[151]

Furthermore, to better meet the requirements of XReco's use cases and pilots, an XReco LICENSE could be developed for the management of non-free rights and for specific clauses or restrictions according to end-users needs.

The XReco platform, by allowing users to import *their own assets*, use *assets from broadcasters* or *from external sources*, must have well-defined procedures for licensing them. Users could also include restrictions on the creation and distribution of assets derivative of other assets and created by means of the XReco services.

[150] https://creativecommons.org/share-your-work/cclicenses/

[151] Benjamin M., et al. "Towards standardization of data licenses: The montreal data license." https://arxiv.org/abs/1903.12262 (2019)

More specifically, for assets ingested into the platform, a license for their use must be specified. For each asset registered on the XReco platform an SLC is created (starting from the template of the chosen licence) which will subsequently notarised on the Blockchain. As reported in Section **Fehler! Verweisquelle konnte nicht gefunden werden.**, there are particular cases where an asset may not have a licence (because it was imported from sources that do not offer one), in this case Text and Data Mining (TDM) Exception or Temporary Copy Exception may be exploited to use such assets during the creation flow of new derived assets through specific XRec0 services.

As described above, users will be able to create derivative assets. In this case, the platform must compare the licences of the assets used to create the derivative asset in two different steps: before the asset is derived and after the asset has been derived. In the first case, the step is necessary to verify whether the licences of the chosen assets are compatible with each other and permit the derivation of assets (e.g., if even only one of the chosen assets had the CC-ND licence, derivatives could not be made in any way). In the second case, the verification is performed to propose licences to the user that are compatible with the licences of the assets used to create the derivative.

8.4 Blockchain services

The blockchain is a distributed ledger that enables the creation and management of a digital, immutable, and transparent register of transactions. It consists of a chain of blocks containing data, each linked to the previous one through cryptographic hash functions (Figure 8-9). These blocks are decentralised and distributed among network participants, eliminating the need for a central authority to validate or record transactions.

Figure 8-9 How blocks are chained.

The technology relies on consensus mechanisms, such as proof-of-work or proof-of-stake, to validate and add new transactions to the chain. This decentralized nature ensures that no single entity has control over the entire network, reducing the risk of fraud and manipulation. With the combination of hash functions, decentralised consensus mechanisms, and interconnection between blocks, the integrity of the blocks is preserved, guaranteeing the immutability of the recorded data.

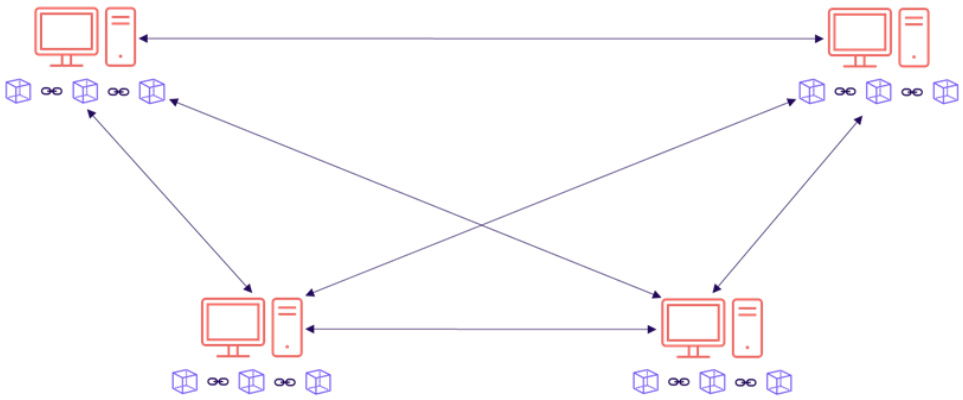


Figure 8-10 Infographic about Blockchain Network.

The blockchain finds application in various areas, particularly in the cryptocurrency realm such as Bitcoin and Ethereum. It acts as the underlying technology for these digital currencies, facilitating secure and transparent peer-to-peer transactions. Furthermore, blockchain has extended its use cases beyond cryptocurrencies to areas such as supply chain, healthcare, finance, and smart contracts.

To notarise SLCs, the Ethereum network will be used. Ethereum is a decentralized, open source blockchain platform that enables the creation and execution of decentralized applications (DApps). A Decentralized Application (or Smart Contract) is a program that runs on the Ethereum blockchain and resides at a specific address.

The primary language used to develop DApps is Solidity. Solidity is a high-level programming language designed specifically for creating smart contracts that runs on the Ethereum Virtual Machine (EVM). Developers use Solidity to define the rules and logic of their smart contracts, which are then compiled into bytecode that can be executed on the Ethereum blockchain and provides features like inheritance, libraries, and complex data structures, making it suitable for building complex applications.

A blockchain network can have two different visibilities:

- Public: (e.g., Bitcoin, Ethereum, Solana, etc) where anyone can access and consult it.
- Private: where only certain entities can access and consult it.

In public blockchain the entire transaction history is visible to all network participants. This transparency is advantageous for use cases requiring trust and decentralization, such as in the creation of decentralized applications and cryptocurrencies. On the other hand, in a private blockchains the access is restricted to a specific group of participants, typically within an organization or consortium. This restriction ensures a higher level of control, privacy, and efficiency.

In both cases, the level of governance and access control can also be selected during the construction of a blockchain:

- Permissioned: the participation is restricted to a predefined set of known entities, requiring explicit permission to join the network. This controlled access ensures a trusted and accountable participant base.
- Permissionless: the participation is not restricted, allowing anyone to join the network without requiring prior authorization.

In a permissioned blockchain, there is a pre-defined set of participants, which provides a sense of control and accountability. This controlled environment is beneficial for scenarios where confidentiality and compliance are paramount, such as in the corporate environment. In contrast, permissionless blockchains prioritise inclusivity, enabling a more decentralised network where trust is established through open participation.

8.4.1 Hyperledger Besu

In XReco, to create a custom blockchain network, we will use Hyperledger Besu. Hyperledger Besu is an Ethereum client designed to be enterprise-friendly for both public and private permissioned network use cases, with an extractable EVM implementation. It can also be run on test networks such as Sepolia and Görli. Hyperledger Besu includes several consensus algorithms including Proof of Stake, Proof of Work, and Proof of Authority (IBFT 2.0, QBFT, and Clique). Its comprehensive permissioning schemes are designed specifically for use in a consortium environment[152].

Hyperledger Besu's utility for creating a custom blockchain network stems from its flexibility in configuration. Organizations can harness Besu to design networks with specific permissions, consensus mechanisms, and privacy settings to suit their unique requirements. Whether it's a private consortium network among trusted partners or a public network requiring open participation, Besu provides the tools to customize the blockchain environment. This adaptability is instrumental for businesses aiming to optimize blockchain technology for their specific needs, ensuring that the network aligns with regulatory standards, privacy considerations, and operational preferences.

8.4.2 Blockchain Service Provider

In the context of XReco, the connection with the blockchain will be made through a microservice called Blockchain Service Provider (BCSP). This service exposes APIs that perform the notarisation of an SLC, its asset revenue and parent assets on the blockchain. Once the transaction is sent over the network, the service responds with the registration id and the hash of the transaction. Using the hash, it is possible to go and verify the validation status of the block it was placed on and to check the notarization over time. To make it easier for the user to check the status of the blockchain (i.e., blocks, transactions, wallets, and smart contracts), there are software tools called Blockchain Explorer, which are discussed in more detail in the next section.

8.4.3 Block Explorer

A Blockchain Explorer (or block explorer) is a software application that allows users to extract, view and review blockchain network metrics. It is a web-based tool that provides a visual representation of the blockchain's transaction history, including details about blocks, transactions, addresses, and network statistics. It enable users to verify transactions, track the flow of cryptocurrency, and monitor the status of the network in real-time.

An example of a block explorer for the Ethererum network is Etherscan (Figure 8-11). However, Etherscan is a closed source project, and you can't use an instance of it for a custom blockchain network.

[152] https://www.hyperledger.org/projects/besu

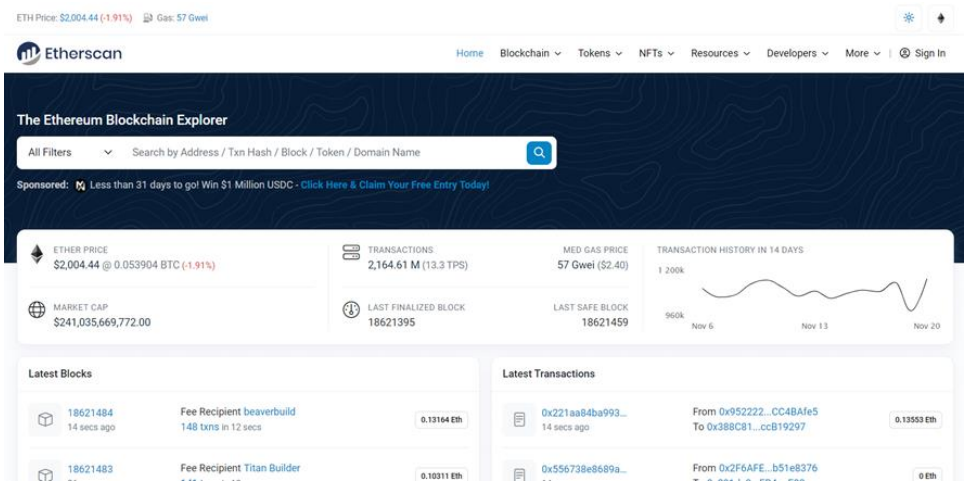


Figure 8-11

Figure 13. Etherscan, the official Blockchain Exploer of the Ethereum Network.

For XReco's blockchain network, we will use an open-source block explorer called Blockscout (Figure 8-12). Blockscout is an open-source alternative to centralized, closed source block explorers such as Etherscan, Etherchain and others[153].

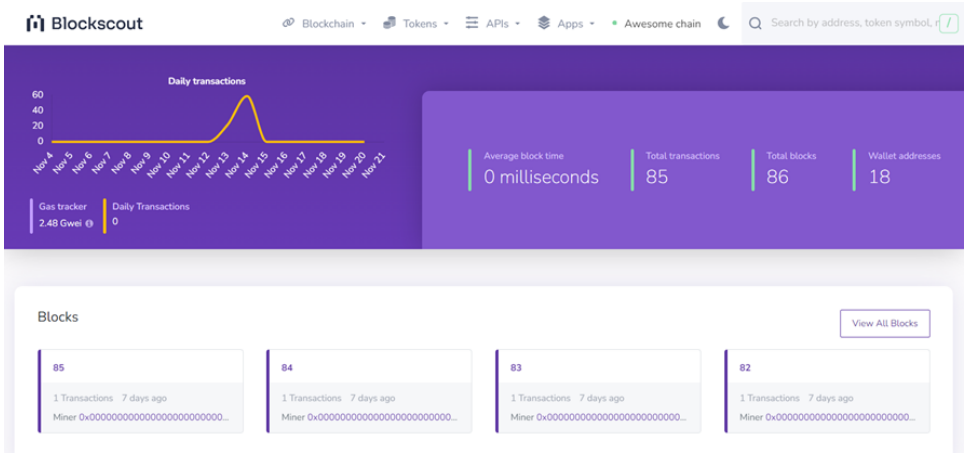


Figure 8-12 Blockscout: Blockchain Explorer for inspecting and analyzing EVM Chains.

Another key point of this software is the ability to verify the SCs code. This functionality allows users to see the contract code and understand better the content of transactions made to interact with the contract itself.

[153] https://github.com/blockscout/blockscout

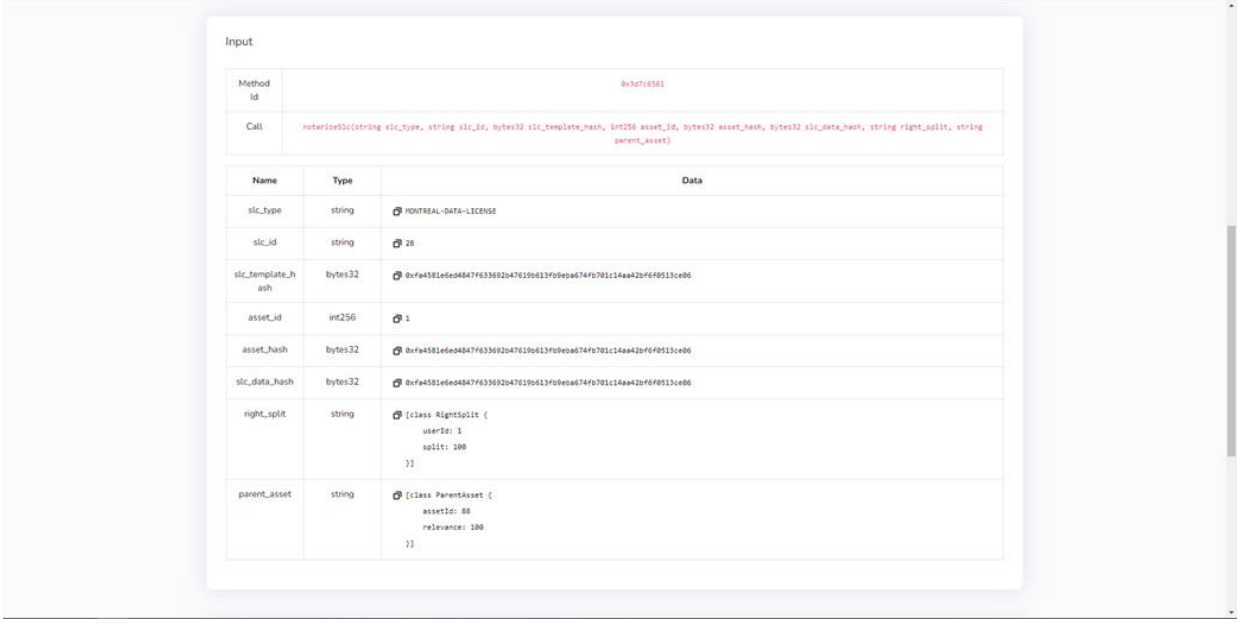


Figure 8-13 Details of a transaction made to a Smart Contract.

As an example, in Figure 8-13 there is the Blockscout page showing the details of an example transaction made by the BCSP. Thanks to the SC verification carried out previously, it's possible to see the call made by the BCSP to a method of the SC and its data. In this case, it is possible to see that the notarised SLC Template is the *Montreal Data License*, the template hash, the hash of the data entered in the template for this particular asset, and so on.

This tool is very important to facilitate the visualisation of what is going on within the Blockchain, but also to verify the history of notarised SLCs, the data of these SLCs and the assets to which they relate, allowing greater exploitation of the strengths of the blockchain: security, transparency, immutability and traceability, in the context of IPR management.

8.5 Data valuation (CERTH)

Data valuation in the context of machine learning refers to the process of associating a specific value with each data sample (or datum) that comprise comprises a model's training dataset, i.e., the process of calculating the importance of each data sample in the final trained machine learning model. Its importance is becoming essential especially in large datasets with data contributed from many parties requiring a compensation depending on the value they provide in the model that they are used on. Data valuation algorithms will be required to be incorporated into most of the services developed in WP4 (more specifically the services that are trained on aggregated datasets coming from many different contributors, i.e., models developed in T4.2 and T4.3), in order to provide the ability to measure the contribution of each training data sample, therefore, offering a percentage measure for the compensation of each sample's contributor. In the following, we describe algorithms of interest in such a context, that will be evaluated for their utilisation in WP4 considering their robustness in measuring consistency, as well as their average efficiency in terms of processing time added to the original training algorithm.

Early works for calculating the importance of data in a specific model were approaching the issue through leave-one-out methodologies (Cook et al., 1977). This is simply the process of retraining the model each time without a specific datum, and evaluating the model's score, with the aim of signifying its importance. However, despite

the fact that such methods are still used today, an approach like this is computationally expensive, as it requires retraining the model multiple times (as many times as the number of samples in the training set). To that end, the utilisation of *influence functions* (a classic technique from robust statistics) was proposed (Koh and Liang, 2017) for identifying which training data are linked with specific predictions of a trained model. While such methods are more efficient in terms of computation time, they were conceived with a slightly divergent contextual framework in mind, i.e., for understanding model behaviour for groups of samples in a training set.

The issue of measuring data contribution can be also addressed through a cooperative game theory context, by the utilisation of Shapley values (Shapley, 1988). In this approach, each datum, can be considered a player in a game, in which the payoff is determined by the average marginal contribution of the player calculated in each of the subsets that the data can form. Shapley values can therefore associate each data sample with their contribution in a trained model. This method and its generalisation (Kwon & Zou, 2021) is the basis for many works in the domain (Ghorbani & Zou, 2019), (Jia et al., 2019), (Schoch et al., 2022). However, the calculation time in Shapley value-based approaches, is exponential increased with the number of samples (or group of samples) present in the dataset. Therefore, many works are utilising Shapley value approximations, either based on Monte Carlo sampling (Maleki et al., 2013), (Ghorbani & Zou, 2019), or by employing learning techniques, such as reinforcement learning (Yoon et al., 2020).

Data valuation via Shapley value-based methods are highly pertinent in the context of XReco. They prove to be a robust measure for calculating the importance of a training set's sample in the final trained algorithm. More specifically, NeRF (T4.2), and 3D reconstruction (T4.3) services, in which neural networks are trained on aggregated data coming from many different contributors, will need to integrate such methods for the requirements of T3.4. However, their computational expense cannot be disregarded, as it reduces user experience. To that end, specific attention will be focused on testing and evaluating more recent approaches, that explicitly target computation time via calculating a data sample's contribution through out-of-bag estimates (Kwon & Zou, 2023) or Data-OOB. Data-OOB is computationally efficient by leveraging trained weak learners and proves to be robust in providing statistical insights into which data are beneficial or detrimental to model training. Data-OOB will be tested and analysed by first incorporating it as a data valuation framework in NeRF training (T4.2), and more specifically in the NeRF-in-the-wild algorithm (described in D4.1). Through its integration, the increase in training time will be evaluated, as well as its influence on the results of the trained NeRF model. Additionally, user feedback in terms of user experience, and actual data-valuation results will be valuable for its adaptation to the XReco services.

9 Outlook (JRS)

At the time of submission of this deliverable, the first set of WP3 results are already available as services, deployable as Docker containers. The focus is thus on completing an integration of these services, in order to enable the first pilots for the use cases.

In parallel to these more development-oriented activities, the research on WP3 components will continue. For the NMR this will include the integration of further description services and completing the support for 3D models throughout the backend, as well as enabling the mixed reality search functionalities envisioned in WP4. In the area of user and workflow management, the implementation of content baskets will be completed in order to use them as means to exchange data between different types of services. Given the decision to support a hosted deployment model, also a local repository component, serving as a per user/organisation content space will be implemented.

For the work on content description services this includes further research on compact learned descriptors. Also further work on few-shot learning will be performed, aiming to address also fuzzier concepts than landmarks, which may be relevant for mining content (e.g., content related to a square, a city quarter). In order to validate these methods on more realistic datasets, an effort for collecting a multimodal landmark dataset has started.

Review and further research under the light of changes in the law, modification or update of technical/technological descriptions and business model choices. Further research exploring the interplay between legal framework and monetization components (derivative output, significance of contributions and smart legal contracts).

The work on rights management and licensing tool will be further driven by the insights from legal analysis (and the emerging legislation). The work on smart legal contracts will further need to be aligned with the work on viable business and licensing models in WP6. The work on data valuation will implement technical components to be integrated with the WP4 reconstruction services and providing input to the licensing tools.

10 References

Aharon, N., Orfaig, R., & Bobrovsky, B. Z. (2022). BoT-SORT: Robust associations multi-pedestrian tracking. arXiv preprint arXiv:2206.14651.

Bailer, Werner, Stefanie Wechtitsch, and Marcus Thaler. "Compressing visual descriptors of image sequences." MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4-6, 2017, Proceedings, Part II 23. Springer International Publishing, 2017.

Boiarov, A., Tyantov, E.: Large scale landmark recognition via deep metric learning. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. pp. 169–178 (2019)

Bonadio Enrico & McDonagh Luke, 'Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity", 2 [2020] Intellectual Property Quarterly, available at: http://eprints.lse.ac.uk/105272/1/Bonadio_McDonagh_IPQ_2020.pdf

Borghi M./S. Karapapa, Non-display uses of copyright works: Google Books and beyond, Queen Mary Journal of Intellectual Property, Vol. 1 No. 1, April 2011, pp. 21–52

Bridy Annemarie, 'Coding Creativity: Copyright and the Artificially Intelligent Author' 5 [2012] Stanford Technology Law Review, available at: https://osf.io/p4feh/download

Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M.: The balanced accuracy and its posterior distribution. In: 2010 20th international conference on pattern recognition. pp. 3121–3124. IEEE (2010)

Caimotti, E., Montagnuolo, M., Messina, A.: An efficient visual search engine for cultural broadcast archives. In: AI* CH@ AI* IA. pp. 1–8 (2017)

Chen, X., Gupta, A.: Webly supervised learning of convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (December 2015)

Chiou Theodoros, Copyright law and algorithmic creativity: Monopolizing inspiration? σε Synodinou, T., Jougleux, P., Markou, C., Prastitou-Merdi, T. (Eds.), EU Internet Law in the Digital Single Market, (Springer 2021), p. 265 ff.

Chiou Theodoros, Copyright lessons on Machine Learning: what impact on algorithmic art?, JIPITEC, 10 (2020), p. 405

Chiou Theodoros, Copyright ownership challenge arising from AI-generated works of art: A time to stand and stare, in I. Stamatoudi (ed.), Research Handbook on Intellectual Property and Cultural Heritage, (EE 2022), pp. 250 ff.

Chiou Theodoros, Non-monetary author's remuneration under EU Copyright Law: the case of data, REDA Conference, [conference presentation] 5 May 2023, Nicosia

Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. Technometrics, 19(1), 15. https://doi.org/10.2307/1268249

Duan, L.Y., Lou, Y., Bai, Y., Huang, T., Gao, W., Chandrasekhar, V., Lin, J., Wang, S., Kot, A.C.: Compact descriptors for video analysis: The emerging mpeg standard. IEEE MultiMedia 26(2), 44–54 (2019). https://doi.org/10.1109/MMUL.2018.2873844Mining Landmark Images from Weakly Annotated Video 13

Erven, T. and Harremos, P. and Shawe-Taylor, J.: Rényi Divergence and Kullback-Leibler Divergence. IEEE Transactions on Information Theory 60(7), pp. 3797 – 3820. https://doi.org/10.1109/TIT.2014.2320500, (2014)

European Economic and Social Committee, Opinion. Action plan/intellectual property, INT/927-EESC-2020-5871

Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24(6), 381–395 (jun 1981). https://doi.org/10.1145/358669.358692, https://doi.org/10.1145/358669.358692

Gasser, R., Arnold, R., Faber, F., & Schuldt, H., Waltenspül, R. Rossetto, L. (2024, January). A new Retrieval Engine for vitrivr. International Conference on Multimedia Modeling (accepted for publication)

Gasser, R., Rossetto, L., Heller, S., & Schuldt, H. (2020, October). Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 4465-4468).

Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

Geiger Christophe, Schönherr Franciska, Stamatoudi Irini, Torremans Paul and Karapapa Stavroula, The Information Society Directive in I. Stamatoudi/Paul Torremans (ed.), EU Copyright Law. A Commentary, 2nd Edition (EE 2021), p. 295

Geiger, Christophe and Iaia, Vincenzo, The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI (October 6, 2023). Computer Law & Security Review, vol 52 (Forthcoming, 2024)

Geiger, Christophe, Statutory Licenses as Enabler of Creative Uses (December 19, 2015). In: R.M. Hilty and K.-C. Liu (eds.), Remuneration of Copyright Owners, Berlin/Heidelberg, Springer, 2017, pp. 305-327, Max Planck Institute for Innovation & Competition Research Paper No. 15-14, Available at SSRN: https://ssrn.com/abstract=2701862

Gervais Daniel, 'The Machine as Author', 105 [2019] Iowa Law Review, available at: SSRN: https://ssrn.com/abstract=3359524

Gezawa, A., Zhang, Y., Wang, Q., Yunqi, L.: A Review on Deep Learning Approaches for 3D Data Representations in Retrieval and Classifications, IEEE Access, https://doi.org/https://doi.org/10.1109/ACCESS.2020.2982196, (2020).

Ghorbani, A., & Zou, J. (2019, May). Data shapley: Equitable valuation of data for machine learning. In International conference on machine learning (pp. 2242-2251). PMLR.

Ginsburg Jane & Budiardjo Luke Ali, 'Authors and Machines', 34 [2019] Berkeley Tech. L. J. 343

Gösgens, M., Zhiyanov, A., Tikhonov, A., Prokhorenkova, L.: Good classification measures and how to find them. Advances in Neural Information Processing Systems 34, 17136–17147 (2021)

Hanocka, R., Hertz, A., Fish, N., Giryes, R., Fleishman, S., Cohen-Or, D.: MeshCNN: a network with an edge. ACM Transactions on Graphics, vol. 38, 1-12, https://doi.org/https://doi.org/10.1145/3306346.3322959, (2019).

Hartmann Christian &. Allan Jacqueline E. M & Hugenholtz Bernt & Quintais João P. & Gervais Daniel, 'Trends and Developments in Artificial Intelligence Challenges to the Intellectual Property Rights Framework. Final

report,' Study Commissioned by the European Commission, prepared by The Joint Institute for Innovation Policy & University of Amsterdam [2020], available at: https://www.ivir.nl/publicaties/download/Trends_and_Developments_in_Artificial_Intelligence-1.pdf

Hugenholtz Bernt and Quintais Joao Pedro, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 International Review of Intellectual Property and Competition Law 1190 sq

Jia, R., Dao, D., Wang, B., Hubis, F. A., Hynes, N., Gürel, N. M., ... & Spanos, C. J. (2019, April). Towards efficient data valuation based on the shapley value. In The 22nd International Conference on Artificial Intelligence and Statistics (pp. 1167-1176). PMLR.

Jing, L., Vahdani, E., Tan, J., Tian, Y.: Cross-Modal Center Loss for 3D Cross-Modal Retrieval, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3142-3151, openaccess.thecvf.com, (2021).

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.

Koh, P. W., & Liang, P. (2017, July). Understanding black-box predictions via influence functions. In International conference on machine learning (pp. 1885-1894). PMLR.

Köklü Kaya, Individual Licensing of Copyrighted Works, In: R.M. Hilty and K.-C. Liu (eds.), Remuneration of Copyright Owners, Berlin/Heidelberg (Springer 2017), pp. 177 ff.

Kwon, Y. & Zou, J.. (2022). Beta Shapley: a Unified and Noise-reduced Data Valuation Framework for Machine Learning . Proceedings of The 25th International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research.

Kwon, Y. & Zou, J.. (2023). Data-OOB: Out-of-bag Estimate as a Simple and Efficient Data Value. Proceedings of the 40th International Conference on Machine Learning, in Proceedings of Machine Learning Research.

Li, C., Deng, C., Li, N., Liu, W., Gao, X., Tao, D.: Self-Supervised Adversarial Hashing Networks for Cross-Modal Retrieval, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 28, pp. 102-112, (2018).

Li, K., Zhang, Z., Wu, G., Xiong, X., Lee, C., Lu, Z., Fu, Y., Pfister, T.: Learning from weakly-labeled web videos via exploring sub-concepts. CoRR abs/2101.03713 (2021), https://arxiv.org/abs/2101.03713

Lin, D., Li, Y., Cheng, Y., Prasad, S., Nwe, T.L., Dong, S., Guo, A: Multi-view 3d object retrieval leveraging the aggregation of view and instance attentive features, Knowl. Based Syst., vol. 247, https://doi.org/https://doi.org/10.1016/j.knosys.2022.108754, (2022).

Lin, M., Ji, R., Liu, H., Wu, Y.: Supervised Online Hashing via Hadamard Codebook Learning, ACM international conference on Multimedia, (2018).

Liu, X., Hu, Z., Ling, H., Cheung, Y.-m.: MTFH: A matrix tri-factorization hashing framework for efficient cross-modal retrieval, IEEE Transactions on Image Processing, vol. 28, (2018).

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)

Lucchi N., ChatGPT: A Case Study on Copyright Challenges for Generative AI Systems (June 12, 2023). European Journal of Risk Regulation (2023), 1–23, Available at SSRN: https://ssrn.com/abstract=4483390 or http://dx.doi.org/10.2139/ssrn.4483390, p. 3, ftnote 12.

Maleki, S., Tran-Thanh, L., Hines, G., Rahwan, T., & Rogers, A. (2013). Bounding the estimation error of sampling-based Shapley value approximation. arXiv preprint arXiv:1306.4265.

Mandal, D., Chaudhury, K. N., Biswas, S.: Generalizes Semantic Preserving Hashing for N-Label Cross-Modal Retrieval, IEEE Transactions on Image Processing, vol. 28, (2018).

Mezuman, E., Weiss, Y.: Learning about canonical views from internet image collections. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (eds.) Advances in Neural Information Processing Systems. vol. 25. Curran Associates, Inc. (2012)

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: European Conference on Computer Vision. pp. 405–421. Springer (2020)

Mittal R., Mechanisms to Make End-Users of Copyrighted Works Pay Through Levy and DRM in Kung-Chung Liu/ Reto M. Hilty, (ed.), Remuneration of copyright owners : regulatory challenges of new business models, (Springer 2017)

Morere, Olivier, et al. "Nested invariance pooling and RBM hashing for image instance retrieval." Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval. 2017.

Noh, H., Araujo, A., Sim, J., Weyand, T., Han, B.: Large-scale image retrieval with attentive deep local features. In: Proceedings of the IEEE international conference on computer vision. pp. 3456–3465 (2017)

Pegia, M., Jónsson, B. Þ., Moumtzidou, A, Dimplaris, S., Gialampoukidis, I., Vrochidis, S., Kompatsiaris, I.: Multimdoal 3D Object Retrieval, International Conference on Multimedia Modeling (MMM2024), (accepted for publication), (2024)

Pegia, M., Jónsson, B. Þ., Moumtzidou, A, Gialampoukidis, I., Vrochidis, S., Kompatsiaris, I.: MuseHash: Supervised Bayesian Hashing for Multimodal Image Representation, ACM International Conference on Multimedia Retrieval (ICMR), https://doi.org/10.1145/3591106.3592228, (2023)

Perd'och, M., Chum, O., Matas, J.: Efficient representation of local geometry for large scale object retrieval. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 9–16 (2009). https://doi.org/10.1109/CVPR.2009.5206529

Perronnin, F., Liu, Y., Renders, J.M.: A family of contextual measures of similarity between distributions with application to image retrieval. In: 2009 IEEE Conference on computer vision and pattern recognition. pp. 2358–2365. IEEE (2009)

Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: Improving particular object retrieval in large scale image databases. In: 2008 IEEE conference on computer vision and pattern recognition. pp. 1–8. IEEE (2008)

Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: 2007 IEEE conference on computer vision and pattern recognition. pp. 1–8. IEEE (2007)

Qi, C.R., H.S., Kaichun, M., Leonidas, J. G.: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), openaccess.thecvf.com, (2017).

Razali, M.N.B., Tony, E.O.N., Ibrahim, A.A.A., Hanapi, R., Iswandono, Z.: Landmark recognition model for smart tourism using lightweight deep learning and linear discriminant analysis. International Journal of Advanced Computer Science and Applications (2023), https://api.semanticscholar.org/CorpusID:257386803

Rosati E., Copyright in the Digital Single Market. Article-by-Article Commentary to the Provisions of Directive 2019/790, Oxford University Press, 2021

Rossetto, L., Giangreco, I., Tanase, C., & Schuldt, H. (2016, October). vitrivr: A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections. In Proceedings of the 24th ACM international conference on Multimedia (pp. 1183-1186).

Rossetto, L., Schuldt, H., Awad, G., Butt, A.A.: V3c–a research video collection. In: MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8–11, 2019, Proceedings, Part I 25. pp. 349–360. Springer (2019)

Samoili Sofia & Lopez Cobo Montserat & Gomez Gutierrez Emilia & De Prato Giuditta & Martinez-Plumed Fernando & Delipetrev Blagoj, 'AI WATCH. Defining Artificial Intelligence', [2020] EUR 30117 EN, Publications Office of the European Union, available at: https://publications.jrc.ec.europa.eu/repository/handle/JRC118163?mode=full

Schoch, S., Xu, H., & Ji, Y. (2022). CS-Shapley: Class-wise Shapley Values for Data Valuation in Classification. Advances in Neural Information Processing Systems, 35, 34574-34585.

Schönberger D., 'Deep Copyright: Up - And Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML)' [2018] SSRN: https://ssrn.com/abstract=3098315

Selvaraju, P., Nabail, M., Loizou, M., Maslioukova, M., Averkiou, M., Andreou, A., Chaudhuri, S., Kalogerakis, E.: BuildingNet: Learning to Label 3D Buildings, IEEE/CVF International Conference on Computer Vision (ICCV), openaccess.thecvf.com, (2021).

Senftleben, Martin and Buijtelaar, Laurens, Robot Creativity: An Incentive-Based Neighboring Rights Approach (October 1, 2020). Available at SSRN: https://ssrn.com/abstract=3707741 or http://dx.doi.org/10.2139/ssrn.3707741, p. 8.

Shapley, L. S. (1953). A value for n-person games.

Song, H., Kim, M., Park, D., Shin, Y., Lee, J.G.: Learning from noisy labels with deep neural networks: A survey. IEEE Transactions on Neural Networks and Learning Systems (2022)

Stamatoudi I. & Torremans P., The Digital Single Market Directive in I. Stamatoudi/Paul Torremans (ed.), EU Copyright Law. A Commentary, 2nd Edition, (EE 2021)

Strowel A.&Vanbrabant B., Copyright licensing: a European view in: J. de Werra (ed.), Research Handbook on IP Licensing, (Edward Elgar Editors 2013), pp. 29-53

Su, H., Maji, S., Kalogerakis, E., Learnined-Miller, E.: Multi-view convolutional neural networks for 3d shape recognition, IEEE Internation Conference on Computer Vision (ICCV), 945-953, cv-foundation.org, (2015)

Su, J.-C., Gadelha, M., Wang, R., Maji, S.: A deeper look at 3d shape classifiers. European Conference on Computer Vision (ECCV), openaccess.thecvf.com, (2018).

Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)

Torii, A., Arandjelovi ́c, R., Sivic, J., Okutomi, M., Pajdla, T.: 24/7 place recognition by view synthesis. In: CVPR (2015)

Torii, A., Sivic, J., Pajdla, T., Okutomi, M.: Visual place recognition with repetitive structures. In: Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 37, pp. 883–890 (06 2013). https://doi.org/10.1109/CVPR.2013.119

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2021). Scaled-yolov4: Scaling cross stage partial network. In Proceedings of the IEEE/cvf conference on computer vision and pattern recognition (pp. 13029-13038).

Wang, X.; Huang, T.; Gonzalez, J.; Darrell, T.; Yu, F. Frustratingly Simple Few-Shot Object Detection. In Proceedings of the International Conference on Machine Learning, Virtual, 12–18 July 2020; pp. 9919–9928.

Wang, Y., Sun, Y., Liu, Z., Sarma, S., Bronstein, M., Solomon, J.: Dynamic graphcnn for learning on point clouds, ACM Transactions on Graphics (TOG), vol. 38, 1-12, https://doi.org/https://doi.org/10.1145/3326362, (2019).

Wang, Z.-D., Chen, L., Xin, R. Li, Xu, X.-S. Xu: Fast Cross-Modal Hashing with Global and Local Similarity Embedding, IEEE Transactions on Cybernetics, (2021).

Weyand, T., Araujo, A., Cao, B., Sim, J.: Google landmarks dataset v2-a large-scale benchmark for instance-level recognition and retrieval. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2575–2584 (2020)

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D ShapeNets: A Deep Representation for Volumetric Shapes, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1912-1920, cv-foundation.org, (2015).

Xie, Y., Liu, Y., Wang, Y., Gao, L., Wang, P., Zhou, K.: Label-attended hashing for multi-label image retrieval, IEEE Transactions on Cybernetics, https://doi.org/https://doi.org/10.1109/TCYB.2021.3059886, (2020).

Yang, M., Cui, C., Xue, X., Ren, H., Wei, K.: 2nd place solution to google landmark retrieval 2020 (2022)

Yokoo, S., Ozaki, K., Simo-Serra, E., Iizuka, S.: Two-stage discriminative reranking for large-scale landmark retrieval. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 1012–1013 (2020)

Yoon, J., Arik, S., & Pfister, T. (2020, November). Data valuation using reinforcement learning. In International Conference on Machine Learning (pp. 10842-10851). PMLR.

Yu, F.: Bdd100k: A large-scale diverse driving video database. BAIR (Berkeley Artificial Intelligence Research). [Online]. Available: https://bair. berkeley.edu/blog/2018/05/30/bdd (2018)

Zhan, W., Wang, Y., Sun, Y., Wu, X.-M., Luo, X., Xu, X.-S.: Discrete online cross-modal hashing, in Pattern Recognition, vol. 122, (2022).

Zhuang, P., Wang, Y., Qiao, Y.: Learning attentive pairwise interaction for fine-grained classification. CoRR abs/2002.10191 (2020), https://arxiv.org/abs/2002.10191

Annex I: Sequence diagrams

This annex presents sequence diagrams related to the intended workflows using the services developed in WP3.

Figure 0-1 provides an overview of the ingest workflow, I.e. registering a new item, importing basic metadata and running content description services. This workflow may be triggered from direct content upload via the orchestrator, from importing an external search result from metasearch or from importing an asset created by one of the WP4 services.

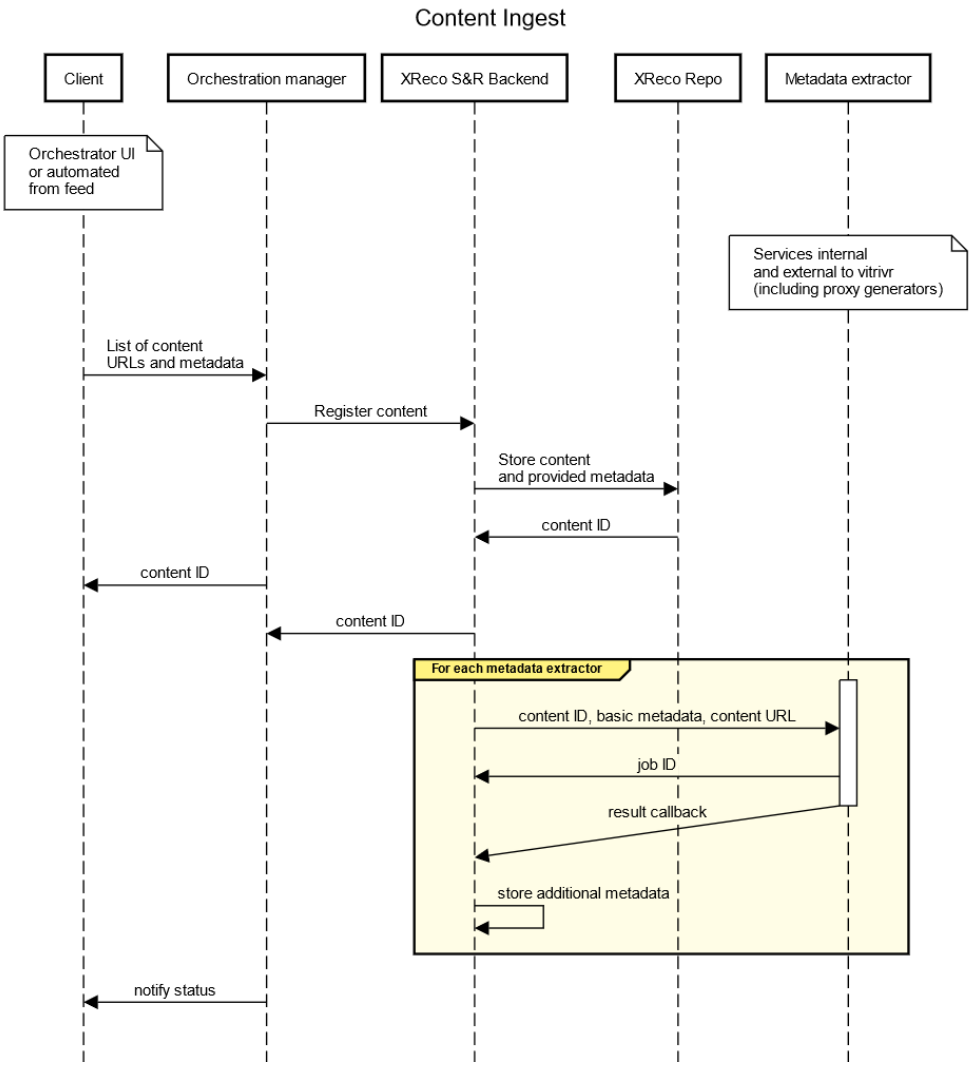


Figure 0-1 Ingest workflow.

Figure 0-2 and Figure 0-3 provide the workflows to local search and metasearch.

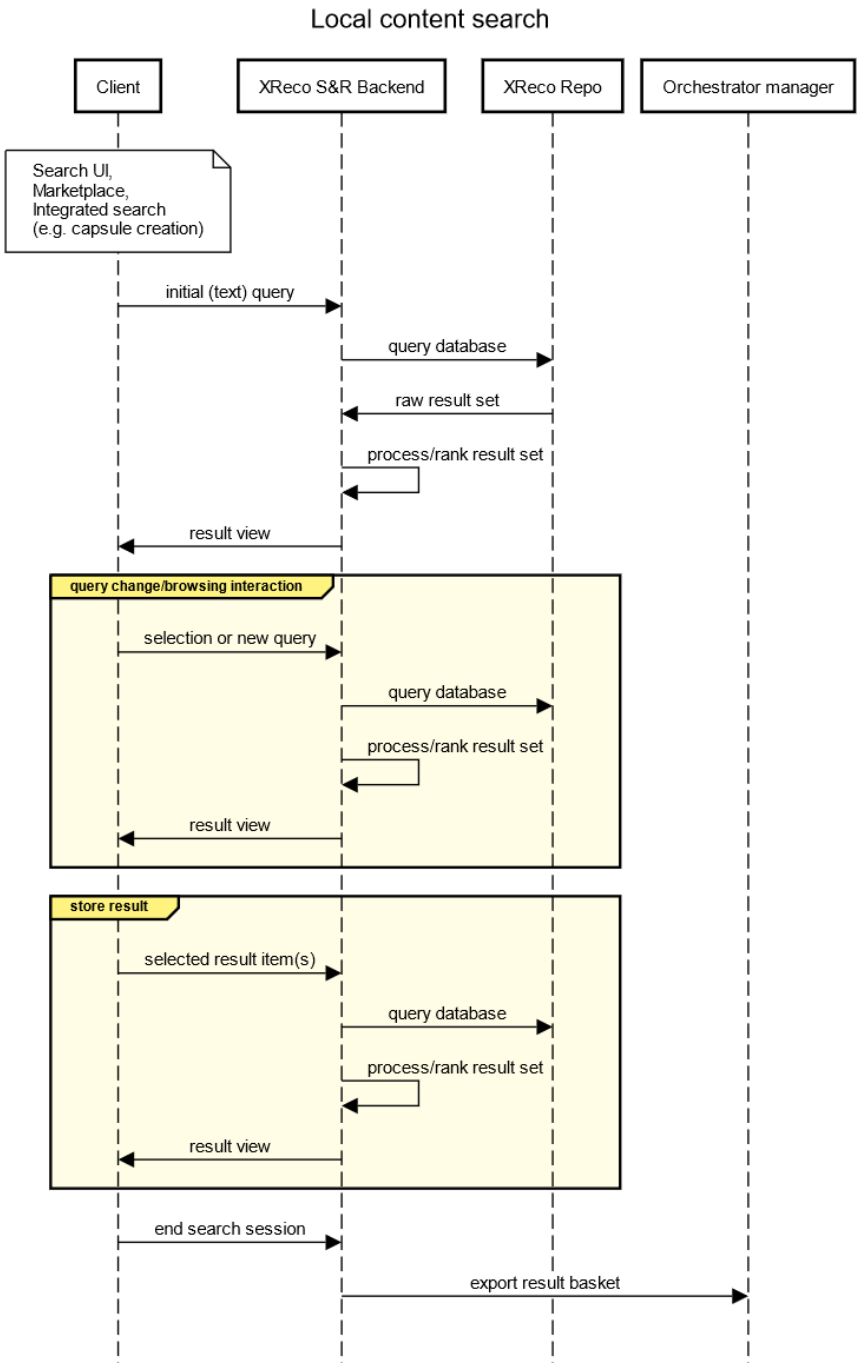


Figure 0-2 Local search workflow.

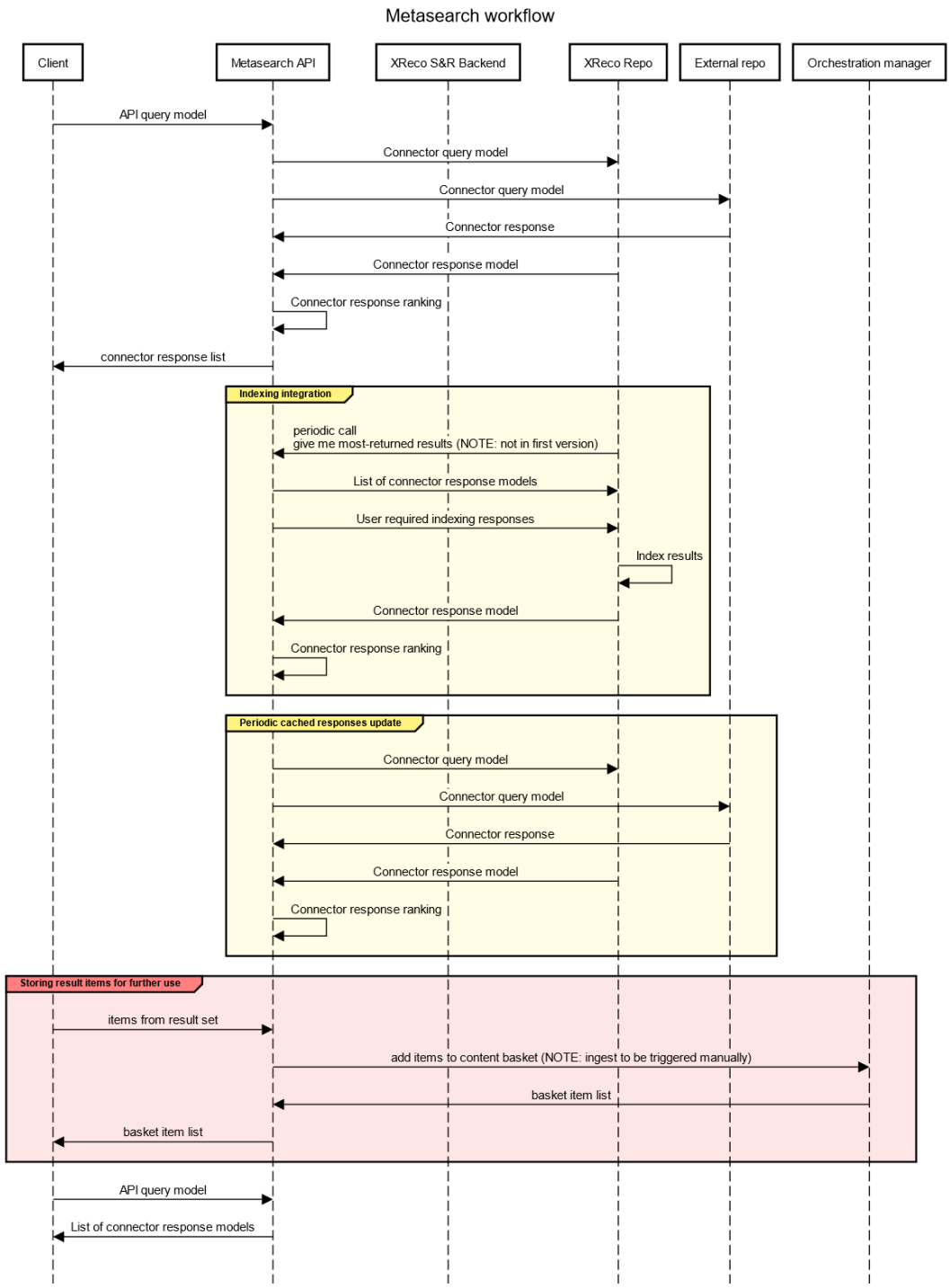


Figure 0-3 Metasearch workflow.

About the position of T3.2 “Cross-modal descriptors” in the XReco project, two implementation options exist (Figure 0-3). First, the user will make a query through a search web app and this app will call the cross retrieval service. The cross retrieval service will call separately the descriptors for each modality (image, text, 3D data) and return the top-k similar results of each modality. Then these results will be fused and visualized to the search web app. Another option is the user will make a query through a search web app and this app will call the cross-retrieval service descriptors. The service will return the fused top-k results from all modalities and return them for visualization to the search web app.

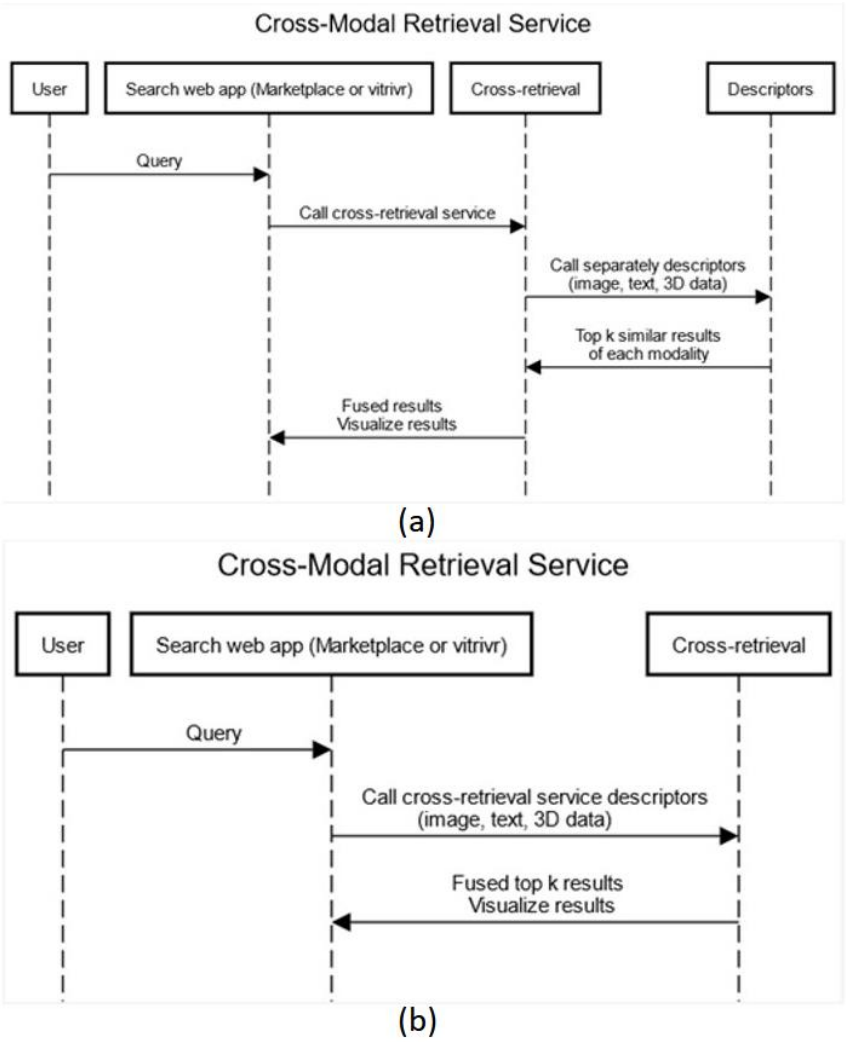


Figure 0-4 Proposed architectures for T3.2 "Cross-modal descriptors".

Figure 0-4 illustrates the architecture of 3D Object retrieval sequence in XReco project. Specifically, the user define a query through a search web app and the Content Basket and then, the app calls the 3D Object Retrieval service for searching relevant object to the given query 3D object. The 3D Object Retrieval service returns the most similar results for visualization in the search web app. In the current version, we put a limit to the returned results in order to evaluate the quality of the first returned 3D objects from the service.

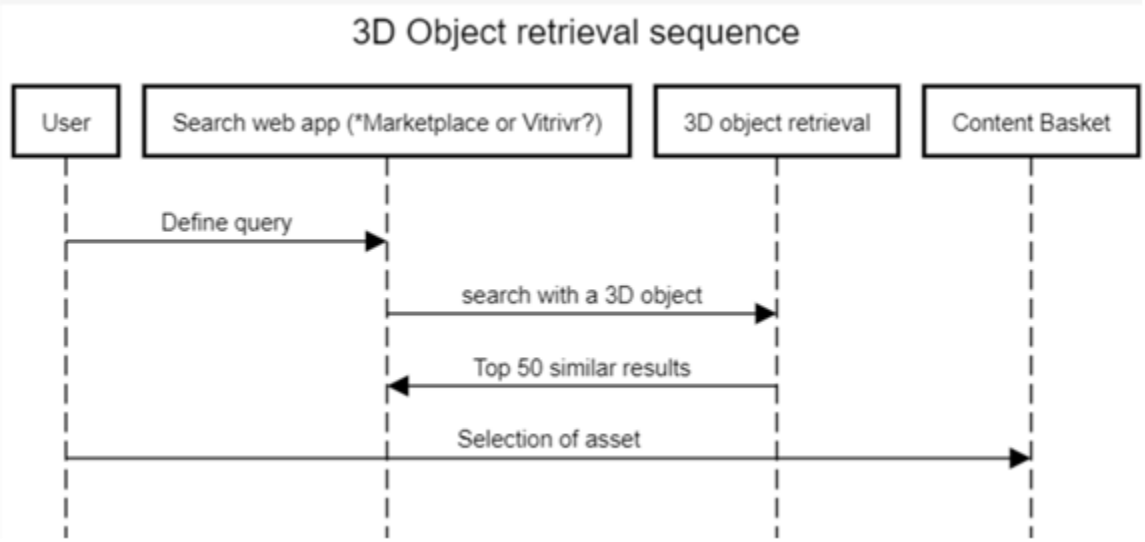


Figure 0-5 Architecture for 3D object retrieval sequence.

11 Annex II: API Definition

The metasearch interactive documentation endpoint is provided at: https://xreco.ari-imet.eu/api/docs

The content description services implement the same API definition, only the job parameters and the result payloads differ. The API definition is provided below.


```
{
    "openapi": "3.0.0",
    "servers": [
        {
            "description": "Local development server",
            "url": "http://localhost:8888"
        }
    ],
    "info": {
        "description": "RESTful definition for communication between vitrivr-engine and an
external feature service",
        "title": "External Feature Extraction Endpoint Definition",
        "version": "1.0.0"
    },
    "paths": {
        "/change_settings": {
            "post": {
                "operationId": "change_settings_change_settings_post",
                "requestBody": {
                    "content": {
                        "application/json": {
                            "schema": {
                                "type": "object"
                            }
                        }
                    },
                    "required": true
                },
                "responses": {
                    "200": {
                        "content": {
                            "application/json": {
                                "schema": {
                                    "type": "string"
                                }
                            }
                        },
                        "description": "Successful Response"
                    },
                    "422": {
                        "content": {
                            "application/json": {
                                "schema": {
                                    "$ref": "#/components/schemas/HTTPValidationError"
                                }
                            }
                        },
                        "description": "Validation Error"
                    }
                },
                "summary": "Change Settings"
            }
        },
        "/extract/{featureName}": {
            "post": {
                "operationId": "extract",
```

```
                "parameters": [
                    {
                        "description": "The name of the feature to be extracted",
                        "in": "path",
                        "name": "featureName",
                        "required": true,
                        "schema": {
                            "type": "string"
                        }
                    }
                ],
                "requestBody": {
                    "content": {
                        "application/json": {
                            "schema": {
                                                    "$ref": "#/components/schemas/Data"
                            }
                        }
                    },
                    "required": true
                },
                "responses": {
                    "200": {
                        "content": {
                            "application/json": {
                                "schema": {
                                    "properties": {
                                        "jobId": {
                                            "anyOf": [
                                                {
                                                    "type": "string"
                                                },
                                                {
                                                    "type": "null"
                                                }
                                            ],
                                            "description":  "The  assigned  jobId  of  the
extraction."
                                        },
                                        "result": {
                                            "anyOf": [
                                                {
                                                    "items": {
                                                        "$ref":
"#/components/schemas/Feature"
                                                    },
                                                    "type": "array"
                                                },
                                                {
                                                    "items": {
                                                        "type": "number"
                                                    },
                                                    "type": "array"
                                                },
                                                {
                                                    "type": "null"
                                                }
                                            ],
                                            "description": "The float vector result or list
of Feature objects"
                                        }
                                    },
                                    "type": "object"
                                }
                            }
                        },
```

```
                        "description": "Successful response"
                    },
                    "404": {
                        "content": {
                            "application/json": {
                                "schema": {
                                                                "$ref":
"#/components/schemas/ErrorStatus"
                                }
                            }
                        },
                        "description": "Not Found"
                    },
                    "422": {
                        "content": {
                            "application/json": {
                                "schema": {
                                    "$ref": "#/components/schemas/HTTPValidationError"
                                }
                            }
                        },
                        "description": "Validation Error"
                    },
                    "500": {
                        "content": {
                            "application/json": {
                                "schema": {
                                                                "$ref":
"#/components/schemas/ErrorStatus"
                                }
                            }
                        },
                        "description": "Internal Server Error"
                    },
                    "503": {
                        "content": {
                            "application/json": {
                                "schema": {
                                                                "$ref":
"#/components/schemas/ErrorStatus"
                                }
                            }
                        },
                        "description": "Service Unavailable"
                    }
                },
                "summary": "Creates a new extraction for a given element",
                "tags": [
                    "External"
                ]
            }
        },
        "/extract/{featureName}/{jobId}": {
            "get": {
                "operationId": "getStatusAndResult",
                "parameters": [
                    {
                        "description": "The name of the feature",
                        "in": "path",
                        "name": "featureName",
                        "required": true,
                        "schema": {
                            "type": "string"
                        }
                    },
                    {
```

```
                        "description": "The jobId of the extraction",
                        "in": "path",
                        "name": "jobId",
                        "required": true,
                        "schema": {
                            "type": "string"
                        }
                    }
                ],
                "responses": {
                    "200": {
                        "content": {
                            "application/json": {
                                "schema": {
                                    "additionalProperties": false,
                                    "properties": {
                                        "result": {
                                            "anyOf": [
                                                {
                                                    "items": {
                                                        "$ref":
"#/components/schemas/Feature"
                                                    },
                                                    "type": "array"
                                                },
                                                {
                                                    "items": {
                                                        "type": "number"
                                                    },
                                                    "type": "array"
                                                },
                                                {
                                                    "type": "null"
                                                }
                                            ],
                                            "description": "The float vector result or list
of Feature objects"
                                        },
                                        "status": {
                                            "enum": [
                                                "completed",
                                                "ongoing",
                                                "cancelled"
                                            ],
                                            "type": "string"
                                        }
                                    },
                                    "required": [
                                        "status"
                                    ],
                                    "type": "object"
                                }
                            }
                        },
                        "description": "Successful response"
                    },
                    "404": {
                        "content": {
                            "application/json": {
                                "schema": {
                                                        "$ref":
"#/components/schemas/ErrorStatus"
                                }
                            }
                        },
                        "description": "Not Found"
```

```
                    },
                    "422": {
                        "content": {
                            "application/json": {
                                "schema": {
                                    "$ref": "#/components/schemas/HTTPValidationError"
                                }
                            }
                        },
                        "description": "Validation Error"
                    },
                    "500": {
                        "content": {
                            "application/json": {
                                "schema": {
                                                                    "$ref":
"#/components/schemas/ErrorStatus"
                                }
                            }
                        },
                        "description": "Internal Server Error"
                    },
                    "503": {
                        "content": {
                            "application/json": {
                                "schema": {
                                                                    "$ref":
"#/components/schemas/ErrorStatus"
                                }
                            }
                        },
                        "description": "Service Unavailable"
                    }
                },
                "summary": "Get the status and result of a specific extraction job",
                "tags": [
                    "External"
                ]
            }
        },
        "/get_settings": {
            "get": {
                "operationId": "get_settings_get_settings_get",
                "responses": {
                    "200": {
                        "content": {
                            "application/json": {
                                "schema": {
                                    "type": "object"
                                }
                            }
                        },
                        "description": "Successful Response"
                    }
                },
                "summary": "Get Settings"
            }
        },
        "/status": {
            "get": {
                "operationId": "status_status_get",
                "responses": {
                    "200": {
                        "content": {
                            "application/json": {
                                "schema": {
```

```
                                    "$ref": "#/components/schemas/StatusService"
                                }
                            }
                        },
                        "description": "Successful Response"
                    }
                },
                "summary": "Status"
            }
        }
    },
    "components": {
        "schemas": {
            "Data": {
                "additionalProperties": true,
                "description": "Base64 encoded data to be processed as a data URL",
                "properties": {
                    "data": {
                        "description": "The min.io URL of the object to be processed",
                        "type": "string"
                    },
                    "end": {
                        "description": "The end time of the segment to be processed",
                        "type": "number"
                    },
                    "last": {
                        "description": "Optional parameter indicating that the segment
represents the last frame",
                        "type": "boolean"
                    },
                    "start": {
                        "description": "The start time of the segment to be processed",
                        "type": "number"
                    }
                },
                "required": [
                    "data",
                    "start",
                    "end"
                ],
                "type": "object"
            },
            "Descriptor": {
                "additionalProperties": false,
                "description": "The extracted descriptor. For example, if a vector feature is
extracted of the data, this is a vector of floats.",
                "type": "object"
            },
            "ErrorStatus": {
                "additionalProperties": false,
                "properties": {
                    "code": {
                        "description": "The error code",
                        "type": "integer"
                    },
                    "description": {
                        "description": "A description of the error",
                        "type": "string"
                    }
                },
                "required": [
                    "code",
                    "description"
                ],
                "type": "object"
            },
```

```
"Feature": {
    "properties": {
        "confidence": {
            "type": "number"
        },
        "label": {
            "type": "string"
        }
    },
    "required": [
        "label",
        "confidence"
    ],
    "type": "object"
},
"HTTPValidationError": {
    "properties": {
        "detail": {
            "items": {
                "$ref": "#/components/schemas/ValidationError"
            },
            "type": "array"
        }
    },
    "type": "object"
},
"Status": {
    "additionalProperties": false,
    "description": "The status of an extraction",
    "properties": {
        "status": {
            "enum": [
                "completed",
                "ongoing",
                "cancelled"
            ],
            "type": "string"
        }
    },
    "required": [
        "status"
    ],
    "type": "object"
},
"StatusService": {
    "additionalProperties": false,
    "properties": {
        "code": {
            "description": "The error code",
            "type": "integer"
        },
        "numb_waiting_jobs": {
            "type": "integer"
        },
        "version": {
            "type": "string"
        }
    },
    "required": [
        "code",
        "version",
        "numb_waiting_jobs"
    ],
    "type": "object"
},
"ValidationError": {
```

```
                "properties": {
                    "loc": {
                        "items": {
                            "anyOf": [
                                {
                                    "type": "string"
                                },
                                {
                                    "type": "integer"
                                }
                            ]
                        },
                        "type": "array"
                    },
                    "msg": {
                        "type": "string"
                    },
                    "type": {
                        "type": "string"
                    }
                },
                "required": [
                    "loc",
                    "msg",
                    "type"
                ],
                "type": "object"
            }
        }
    }
}
```

12 Annex III: Datasets

12.1 Pre-existing Datasets used in WP3

Dataset name	MSCOCO
Purpose	WP3, Training and validation of object detection algorithms
Origin	Publicly available dataset
Dataset description	COCO is a large-scale object detection, segmentation, and captioning dataset for object segmentation, recognition in context, and superpixel stuff segmentation. It contains 330K images (>200K labelled), 1.5 million object instances, 80 object categories, 91 stuff categories, 5 captions per image and 250,000 people with key points
URL	https://cocodataset.org
Standards and metadata	Images are in JPEG format; annotations are provided as JSON files describing polygons (the format is widely used for visual analysis tasks)
Data license / sharing	The dataset is published under Creative Commons Attribution 4.0 License.
Personal data	Persons are depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	ModelNet40
Purpose	WP3, Training and validation of retrieval algorithms
Origin	Publicly available dataset
Dataset description	ModelNet40 is a large-scale 3D CAD mode dataset, encompassing a diverse array of object categories, such as cars, bottles, and more. It

	contains 12,311 items, 40 object categories. Each item has one image, mesh and point-cloud view/modality.
URL	https://modelnet.cs.princeton.edu/
Standards and metadata	Images are in PNG format; Mesh in OBJ or OFF format and Point Clouds in PLY format. We extracted representations for each modality and we provided as metadata.
Data license / sharing	The dataset is published under Creative Commons Attribution 4.0 License.
Personal data	None
Storage duration	Undetermined
Access	All members of the project team

Dataset name	BuildingNet_v0
Purpose	WP3, Training and validation of retrieval algorithms
Origin	Publicly available dataset
Dataset description	BuildingNet_v0 is a large-scale dataset of annotated 3D building models whose exteriors and surroundings are consistently labelled. It contains 2000 items, 60 object categories. Each item has one image, mesh and point-cloud view/modality.
URL	https://github.com/buildingnet/buildingnet_dataset
Standards and metadata	Images are in PNG format; Mesh in OBJ or OFF format and Point Clouds in PLY format. We extracted representations for each modality and we provided as metadata.
Data license / sharing	The dataset is published under Creative Commons Attribution 4.0 License.
Personal data	None
Storage duration	Undetermined
Access	All members of the project team

Dataset name	XR4DRAMA
Purpose	WP3, Training and validation of object detection algorithms

Origin	Dataset created for the purposes of XR4DRAMA EU project
Dataset description	The dataset consists of 14 textured 3D models (9 are of larger areas of the town and 5 of individual buildings) created with photogrammetry from videos originated from drone footage and handheld cameras of monuments around the cities of Corfu and Vicenza. The 3D models were created by up2metric (https://www.up2metric.com/) and they are described here: Symeonidis, Spyridon, et al. "An Extended Reality System for Situation Awareness in Flood Management and Media Production Planning." Electronics 12.12 (2023): 2569.
URL	https://xr4drama.eu/ and https://www.up2metric.com/
Standards and metadata	3D models are in OBJ format; textures are in JPEG and MTL format.
Data license / sharing	The dataset is published under Creative Commons Attribution 4.0 License.
Personal data	No persons are depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	Places365
Purpose	WP3, Training and validation of classifiers for types of indoor and outdoor locations
Origin	Publicly available dataset
Dataset description	Places365 is a dataset of images of indoor and outdoor locations, containing 10 million images comprising 400+ unique scene categories The categories are grouped into 16 super categories, which are again grouped into indoor, outdoor/natural and outdoor/manmade.
URL	ttp://places2.csail.mit.edu/download.html

Standards and metadata	Images are in JPEG format, metadata provided as CSV.
Data license / sharing	The license terms are defined as: • You will use the data only for non-commercial research and educational purposes. • You will NOT distribute the above images. • Massachusetts Institute of Technology makes no representations or warranties regarding the data, including but not limited to warranties of non-infringement or fitness for a particular purpose. • You accept full responsibility for your use of the data and shall defend and indemnify Massachusetts Institute of Technology, including its employees, officers and agents, against any and all claims arising from your use of the data, including but not limited to your use of any copies of copyrighted images that you may create from the data.
Personal data	Persons are depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	Google Landmarks v2
Purpose	WP3 and WP4, Training and validation of NeRF and 3D reconstruction algorithms for producing trained models of specific scenes; training and validation of landmark image classification. Only landmark instances with a sufficient number of images and that are relevant to the project's demonstrators will be used.
Origin	Publicly available dataset
Dataset description	Google Landmarks Dataset v2 contains images annotated with labels representing human-made and natural landmarks. The dataset is mainly used for landmark recognition and retrieval research; however, it contains some landmark instances captured from many angles, and at different times of day (showcasing variability in weather, lighting etc) and can be suitable for NeRF training.

URL	https://github.com/cvdfoundation/google-landmark
Standards and metadata	Images are in JPEG format, metadata provided as CSV.
Data license / sharing	All the images in the training-set have CC-BY licenses without the NonDerivs (ND). The license allows uses to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. Specific licenses for each image instance are contained in a single metadata file.
Personal data	Persons are depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	Rafa Volograms
Purpose	WP3, Training and validation of volumetric and video descriptors with retrieval.
Origin	Publicly available dataset
Dataset description	Five second sequence with some dance moves. Rafa was captured at V-SENSE's 12 camera studio in Dublin, Ireland. Meshes are ~40 polys/frame and texture images are 4069x4069.
URL	https://drive.google.com/drive/folders/1K0dzAVV1td3dZH5y9oU_qW11FubCVemD
Standards and metadata	Images are in JPEG format, mesh and texture information.
Data license / sharing	The Dataset is the exclusive property of Volograms Ltd. The licensee may not assign its rights and obligations under this Agreement in whole or in part without the prior written consent of the Licensor. The Licensor may assign any of its rights or obligations under this Agreement without the consent of the licensee being required.

Personal data	Person might be depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	Levi Volograms
Purpose	WP3, Training and validation of volumetric and video descriptors with retrieval.
Origin	Publicly available dataset
Dataset description	Five second dancing sequence featuring Levi, an incredibly talented performer. Levi was captured in a 60-camera studio in the California, US. Meshes are ~40 polys/frame and texture images are 4069x4069.
URL	https://drive.google.com/drive/folders/1NvAUN0Xwp3golDxLRZvGp4C2UNIAYv69
Standards and metadata	Images are in JPEG format, mesh and texture information.
Data license / sharing	The Dataset is the exclusive property of Volograms Ltd. The licensee may not assign its rights and obligations under this Agreement in whole or in part without the prior written consent of the Licensor. The Licensor may assign any of its rights or obligations under this Agreement without the consent of the licensee being required.
Personal data	Person might be depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	Sir Frederick Volograms
Purpose	WP3, Training and validation of volumetric and video descriptors with retrieval.
Origin	Publicly available dataset
Dataset description	One minute monologue sequence. Sir Frederick was captured in a 12-camera studio captured at V-SENSE's 12 camera studio in Dublin, Ireland. Meshes are ~40 polys/frame and texture images are 4069x4069.
URL	https://drive.google.com/drive/folders/1NplcQijc3j5gcpgvDzojVPqAtEQP7p2J
Standards and metadata	Images are in JPEG format, mesh and texture information.
Data license / sharing	The Dataset is the exclusive property of Volograms Ltd. The licensee may not assign its rights and obligations under this Agreement in whole or in part without the prior written consent of the Licensor. The Licensor may assign any of its rights or obligations under this Agreement without the consent of the licensee being required.
Personal data	Person might be depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	Hand Gesture Recognition Database
Purpose	WP3, Training and validation of image descriptors with retrieval.
Origin	Publicly available dataset
Dataset description	The database is composed by 10 different hand-gestures (showed above) that were performed by 10 different subjects (5 men and 5 women). Each subfolder contains 200 images.

URL	https://www.kaggle.com/datasets/gti-upm/leapgestrecog
Standards and metadata	Images are in JPEG format and label annotation is included
Data license / sharing	The Dataset can be used under the CC BY-NC-SA 4.0 licence.
Personal data	Persons might be depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	HaGRID - HAnd Gesture Recognition Image Dataset
Purpose	WP3, Training and validation of image descriptors with retrieval.
Origin	Publicly available dataset
Dataset description	HaGRID size is 716GB and dataset contains 552,992 FullHD (1920 × 1080) RGB images divided into 18 classes of gestures. Also, some images have no_gesture class if there is a second free hand in the frame. This extra class contains 123,589 samples. The data were split into training 92%, and testing 8% sets by subject user_id, with 509,323 images for train and 43,669 images for test.
URL	https://github.com/hukenovs/hagrid
Standards and metadata	Images are in JPEG format and label annotation is included
Data license / sharing	This dataset is licensed under a variant of Creative Commons Attribution-ShareAlike 4.0 International License.
Personal data	Persons might be depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	PATS Dataset (Pose, Audio, Transcript, Style)
Purpose	WP3, Training and validation of text (transcript) to pose descriptors with retrieval.
Origin	Publicly available dataset
Dataset description	PATS was collected to study correlation of co-speech gestures with audio and text signals. The dataset consists of a diverse and large amount of aligned pose, audio and transcripts. With this dataset, we hope to provide a benchmark which would help develop technologies for virtual agents which generate natural and relevant gestures.
URL	https://chahuja.com/pats/
Standards and metadata	Images are in JPEG format and label annotation is included
Data license / sharing	The person may download and use this database only after signing and returning this agreement form. By signing this document, the user agrees to the following terms: 1. Commercial and academic use: the database is made available for research purposes only. Any commercial use of this data is forbidden. 2. Redistribution: the user may not distribute the database or parts of it to any third party. 3. Publications: the use of data for illustrative purposes in publications is allowed. Publications include both scientific papers, and presentations for scientific/educational purposes. 4. Citation: all publications must cite the following related papers for using the dataset: Chaitanya Ahuja, Dong Won Lee, Yukiko I. Nakano, and Louis-Philippe Morency. "Style Transfer for Co-Speech Gesture Animation: A Multi-Speaker Conditional-Mixture Approach." ECCV 2020 Chaitanya Ahuja, Dong Won Lee, Ryo Ishii, and Louis-Philippe Morency. "No gestures left behind: Learning relationships between spoken language and freeform gestures" Findings at EMNLP 2020

	5. EULA changes: the Language Technologies Institute at Carnegie Mellon University is allowed to change these terms of use at any time. In this case, users will be informed of the changes and will have to sign a new agreement form to keep using the database. 6. Warranty: the database comes without any warranty. In no event shall the provider be held responsible for any loss or damage caused by the use of this data.
Personal data	Persons might be depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	RAI Monuments of Italy
Purpose	WP3, Training and validation of landmark classification; testing search and retrieval.
Origin	Provided by RAI under a custom license agreement
Dataset description	The dataset contains about 2,000 clips depicting about 200 monuments from all regions of Italy, mainly acquired from RAI regional newscasts, collected for assessing similarity search in video. Annotations of the monument are provided on clip level. Each clip contains typically a news story, of which one or more shots contain an exterior view of the relevant monument, and in some cases also interior views. Some of the shots may show the monument occluded or in the background (e.g., as backdrop of an interview). In addition, the clips often contain other material of the story, e.g., the anchor in the studio introducing the topic (with an image that shows a view of the monument or something else), views of people in the street, close-up shots of people or interior items etc. As such, the dataset is typical for the type of content and the granularity of annotation to be found in a broadcast archive.
URL	Distributed via project internal SFTP server

Standards and metadata	Videos in MPEG-4 format, organised in directories by monument type. The directory names encode the region and monument.
Data license / sharing	The license permits use of research purposes, within the groups signing the license. Showing excerpts are part of scientific reporting and publications is permitted, other public use of the content requires permission from RAI.
Personal data	Persons are depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

Dataset name	Vimeo Creative Commons Collection (V3C)
Purpose	WP3, Evaluation of search and retrieval
Origin	Created at UNIBAS for use in the TRECVID and VBS benchmarks
Dataset description	V3C is composed of 28,450 videos collected from the video sharing platform Vimeo. Apart from the videos themselves, the collection includes meta and shot-segmentation data for each video, together with the resulting keyframes in original as well as reduced resolution
URL	See Rossetto, Luca, et al. "V3C–a research video collection." MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8–11, 2019, Proceedings, Part I 25. Springer International Publishing, 2019. https://arxiv.org/pdf/1810.04401.pdf For download info see https://www-nlpir.nist.gov/projects/tv2023/data.html
Standards and metadata	Videos in MPEG-4 format, text files with basic video metadata, keyframes in JPEG format.

Data license / sharing	The collections contains only videos published under Creative Commons Attribution (CC-BY) license.
Personal data	Persons are depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

12.2 Datasets created by WP3

Dataset name	Weakly Annotated Video Landmarks (WAVL)
Purpose	WP3, Training and validation of landmark image classification
Origin	Created by JRS in T3.1
Dataset description	In order to mimic a keyframe dataset as extracted from archive video content, we merged images from two different sources: the Google Landmarks v2 dataset and the V3C video dataset. These combined sources allowed us to create a dataset that contains sets of keyframes as they would be extracted from one video, containing both keyframes of a particular landmark as well as unrelated keyframes (noise). The video is annotated with the landmark visible in the subset of keyframes taken from Google Landmarks. For each set of keyframes representing a video, we combined on average 30 associated images from the Google Landmarks v2 dataset with on average 11 keyframes (noise images) from one of the videos in the V3C1 dataset. This process has been done for 141 landmarks, resulting in a dataset of 5,770 images. The dataset release does not republish images from the source datasets, but contains code and annotations to contruct the dataset.
URL	https://github.com/XRecoEU/WAVL-Dataset

Standards and metadata	Python code and text files.
Data license / sharing	Creative Commons Attribution (CC-BY) license (as both of source datasets)
Personal data	Persons are depicted in the image used in dataset.
Storage duration	Undetermined
Access	Public

Dataset name	RAI Drones clips
Purpose	WP3, Creation 3D model and point cloud, training and validation of landmark classification; testing search and retrieval.
Origin	Created and Provided by RAI under a custom license agreement
Dataset description	The dataset created consists of 100 clips captured in various Italian cities and diverse environments, encompassing both urban and natural settings. These clips showcase monuments and natural subjects, emphasizing a comprehensive approach to capturing scenes from all perspectives. The primary goal is to create detailed point cloud and 3D models, and as such, the majority of the clips aim to capture subjects from multiple viewpoints. Each clip, acquired through the use of drones, presents different focal lengths and camera movements to ensure the creation of the most comprehensive 3D models possible. The dataset is carefully curated to include both monuments and natural elements, reflecting a rich variety of scenes. Annotations are provided at the clip level to facilitate evaluation and analysis. This dataset aligns with the objectives of the project, offering a diverse and detailed collection of clips for research and development in the realm of 3D modeling and similarity search in videos.
URL	Distributed via project internal server
Standards and metadata	Videos in MPEG-4 format, organised in directories by city, region or environment type.
Data license / sharing	The license permits use of research purposes, within the groups signing the license. Showing excerpts are part of scientific reporting and

	publications is permitted, other public use of the content requires permission from RAI.
Personal data	Persons are depicted in the dataset.
Storage duration	Undetermined
Access	All members of the project team

13 Annex IV: Supplementary information for Section 4

13.1 Metrics used for evaluating multimodal descriptors

MAP calculates the average precision across all relevant items in the ranked list of results. It considers both the precision and the ranking of relevant items, providing a comprehensive measure of how well the system retrieves relevant results across different queries.

$$mAP = \frac{\sum_{i=1}^{|Q|} AP_q}{|Q|}, AP_q = \frac{(precision@k)_q \times rel(k)}{|R|}$$

(5)

R is the total number of relevant items and rel(k) is an indicator function equal to 1 if the item at rank k is relevant, and 0 otherwise.

Precision@k measures the proportion of relevant items among the top k results. It assesses how well the system performs in terms of accuracy within the top k items. A higher precision@k indicates that a larger proportion of the top results are relevant.

$$precision@k = \frac{Number\ of\ relevant\ items\ in\ top-k}{k}$$

(6)

Recall@k evaluates the proportion of relevant items retrieved among all relevant items in the dataset, up to the top k results. It provides insights into how well the system captures all relevant items within the specified range (up to k). A higher recall@k suggests that the system is effective in retrieving relevant items.

$$recall@k = \frac{Number\ of\ relevant\ items\ in\ top-k}{Total\ number\ of\ relevant\ items}$$

(7)

Fscore@k is the harmonic mean of precision and recall at k. It combines both precision and recall into a single metric, providing a balanced assessment of the system's performance. A higher fscore@k indicates a good balance between precision and recall.

$$fscore@k = \frac{2 \times precision@k \times recall@k}{precision@k + recall@k}$$

(8)

14 Annex V: Data Legal Context

The EU has become a global trailblazer in setting rules for a range of activities concerning data and digital phenomena. Beyond rules concerning data protection and privacy,[154] the EU has adopted legislative measures regarding rules for contracts between traders and consumers for the supply of digital content or digital services,[155] updated rules on copyright and related rights in the Digital Single Market,[156] rules for the Free Flow of Non-Personal Data,[157] rules on "open data" and the re-use of public sector information,[158] rules for data governance,[159] rules for certain digital platforms,[160] with new rules being proposed regarding data and artificial intelligence.[161] Not all of these regulatory instruments address the situations emerging within the context of XReco. Some will be addressed in the specific sections dedicated to particular issues that they do address.[162] The role of some of these new legal regimes will nevertheless be of relevance to the ultimate design of XReco as a platform providing functionalities and tools for interacting with data and content in various ways.

Firstly, rules for the contracts between traders and consumers for the supply of digital content or digital services may play an important role in the digital economy. Such rules have been established by the so-called Digital Content Directive,[163] "laying down common rules on certain requirements concerning contracts between traders and consumers for the supply of digital content or digital services" .[164] Specifically, the Digital Content Directive regulates the contract between a trader – "any natural or legal person, irrespective of whether privately or publicly owned, that is acting, including through any other person acting in that natural or legal person's name or on that person's behalf, for purposes relating to that person's trade, business, craft, or profession, in relation to contracts covered by this Directive"[165] – and consumers – "any natural person who, in relation to contracts covered by this Directive, is acting for purposes which are outside that person's trade, business, craft, or profession".[166] The Directive clarifies that platform providers "could be considered to be traders under this Directive if they act for purposes relating to their own business and as the direct contractual partner of the consumer for the supply of digital content or a digital service".[167] Further, where the contract is concluded "for purposes that are partly within and partly outside the person's trade", EU Member States remain free to

[154]GDPR;Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)[2002] OJ L 201/37 (ePrivacy Directive), see also Commission, 'Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)' COM(2017) 10 final (ePrivacy Regulation Proposal).
[155]Digital Content Directive, see Art. 1.
[156]Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC[2019] OJ L 130/92 (CDSM Directive).
[157]Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union[2018] OJ L 303/59 (FFDR).
[158]Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)[2019] OJ L 172/56 (Open Data Directive).
[159]DGA.
[160]DSA; DMA.
[161]Data Act; Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts' COM(2021) 206 final (AI Act Proposal).
[162]See the section on the applicability of copyright in regard to the CDSM Directive; See the section on smart legal contracts regarding the role of the Data Act.
[163] Digital Content Directive.
[164]Art. 1 Digital Content Directive.
[165]Art. 2(5) Digital Content Directive.
[166]Art. 2(6) Digital Content Directive.
[167]Recital 18 Digital Content Directive.

determine whether and under what conditions that person should be considered a consumer.[168] As XReco is fundamentally a business-to-business platform, it therefore does not qualify as a trader in the context of the Digital Content Directive, as it does not enter into direct contracts with consumers. Further, as potential contracts between XReco participants would be for the purpose of those participant's trade, business, craft or profession (i.e., between "traders"), these contracts fall outside the scope of the Digital Content Directive. XReco participants should nevertheless be cognizant of the Digital Content Directive's rules when engaging in consumer-facing business.

Secondly, the EU has provided a framework for the "free flow of non-personal data" (FFDR).[169] In this regard, the EU acted by "laying down rules relating to data localisation requirements, the availability of data to competent authorities and the porting of data for professional users".[170] The FFDR applies to the "processing of electronic data other than personal data in the Union, which is (a) provided as a service to users residing or having an establishment in the Union, regardless of whether the service provider is established or not in the Union; or (b) carried out by a natural or legal person residing or having an establishment in the Union for its own needs."[171] Whereas processing is defined as "any operation or set of operations which is performed on data or on sets of data in electronic format" .[172] As some have indicated, the FFDR is perhaps most influential in determining "non-personal data" as a legal category of its own.[173] In sum, however, the provisions of the FFDR address the Member States and the Commission,[174] meaning that in spite of this scope of applicability, no provision directly affects the legal position of private sector entities, such as those at stake in the context of XReco.

Thirdly, under the Open Data Directive,[175] the EU has adopted rules "governing the re-use and the practical arrangements for facilitating the re-use of" documents held by public sector bodies of the EU Member States, certain documents held by public undertakings and certain research data.[176] Regarding research data, the scope of application extends beyond public sector entities in certain regards,[177] requiring that "[without prejudice to [Art. 1(2)(c) Open Data Directive], research data shall be re-usable for commercial or non-commercial purposes in accordance with Chapters III and IV, insofar as they are publicly funded and researchers, research performing organisations or research funding organisations have already made them publicly available through an institutional or subject-based repository" .[178] In that regard, publicly-funded research data, where it overlaps with XR content and data, may be relevant in the context of a participants to XReco. Further, the Open Data Directive recognises that due to the existence of "cooperation arrangements between libraries, including university libraries, museums, archives and private partners, which involve digitisation of cultural resources granting exclusive rights to private partners," it may be necessary "to give the private partner the possibility to

[168]Recital 17 Digital Content Directive.
[169] FFDR.
[170]Art. 1 FFDR.
[171]Art. 2(1) FFDR.
[172]Art. 3(2) FFDR.
[173]Thomas Streinz, 'The Evolution of European Data Law' in Craig and de Búrca (eds), *The Evolution of EU Law* (OUP 2021).
[174]Artt. 4, 5 and 7 FFDR address the Member States, Art. 6 FFDR addresses the Commission, whereas Art. 8 FFDR addresses both.
[175] Open Data Directive.
[176]Art. 1(1) Open Data Directive.
[177] "[Pursuant] to the conditions set out in Article 10" (Art. 1(1)(c) Open Data Directive).
[178]Art. 10(2) Open Data Directive.

recoup its investment" relating to the "digitisation of cultural resources".[179] Where such a right is in place, it may interfere with the ability for XR content or data to be generated based on digitalised cultural resources.[180]

Fourthly, the EU has adopted a number of rules addressing the conditions for data sharing under the Data Governance Act (DGA).[181] The reach of the DGA is broad, laying down rules on the "conditions for the re-use, within the Union, of certain categories of data held by public sector bodies",[182] "a notification and supervisory framework for the provision of data intermediation services",[183] "a framework for voluntary registration of entities which collect and process data made available for altruistic purposes",[184] and "a framework for the establishment of a European Data Innovation Board".[185] It should be noted that the rules on the re-use of protected data held by public sector bodies does not apply to certain public media organisations, clarifying that the relevant Chapter of the DGA does not apply to "data held by public service broadcasters and their subsidiaries, and by other bodies or their subsidiaries for the fulfilment of a public service broadcasting remit",[186] nor to "data held by cultural establishments and educational establishments".[187] Unlike the FFDR, therefore, the DGA also directly addresses private sector actors as potential providers of data intermediation services or data altruism organisations. Therein, the definition of what comprises a "data intermediation service" remains unclear,[188] with potentially significant implications for the data economy and the treatment of non-personal data via such services.[189]

Two regulations that have been adopted in the area of digital markets are the Digital Markets Act (DMA) and the Digital Services Act (DSA).[190] Whereas the DMA focuses on the "core platform services" provided by "gatekeepers",[191] which are significantly large entities operating in the internal market,[192] the DSA codifies provisions that limit the liability of mere conduits, caching and hosting service providers,[193] previously regulated at the EU level by the E-Commerce Directive.[194] The DSA also specifies the obligation of providers of

[179]Recital 49 Open Data Directive.
[180]On the tension created by this, particularly in relation to the Copyright in the Digital Single Market Directive, see Andrea Wallace and Ellen Euler, 'Revisiting Access to Cultural Heritage in the Public Domain: EU and International Developments' (2020) 51 IIC - International Review of Intellectual Property and Competition Law 823.
[181] DGA.
[182]Art. 1(1)(a) DGA.
[183]Art. 1(1)(b) DGA.
[184]Art. 1(1)(c) DGA.
[185]Art. 1(1)(d) DGA.
[186]Art. 3(2)(b) DGA.
[187] Art. 3(2)(c) DGA.
[188] Tervel Bobev and others, 'White Paper on the Definition of Data Intermediation Services' (CiTiP White Paper Series 2023, 2 October 2023).
[189] Leander Stähler, 'The Problem of Regulating Data Intermediaries: Insights from the Public Utilities Doctrine' (2024, forthcoming).
[190]DMA; DSA.
[191]Artt. 1 and 2 DMA; which shall be designated as gatekeepers if "(a) it has a significant impact on the internal market; (b) it provides a core platform service which is an important gateway for business users to reach end users; and (c) it enjoys an entrenched and durable position, in its operations, or it is foreseeable that it will enjoy such a position in the near future." (Art. 3(1) DMA).
[192]Art. 3(2) DMA.
[193]Artt. 4-6 DSA.
[194]Artt. 12-14 Directive 2000/31/EC [2000] OJ L 178/1 (E-Commerce Directive).

"intermediary services"[195] to comply with orders to act against illegal content,[196] specifying a suite of general and specific due diligence obligations for online content.[197]

Proposed on 23 February 2022, the Data Act (DA) would address numerous practices in regard to data.[198] Specifically, the DA would provide "harmonised rules on making data generated by the use of a product or related service available to the user of that product or service, on the making data available by data holders to data recipients, and on the making data available by data holders to public sector bodies or Union institutions, agencies or bodies, where there is an exceptional need, for the performance of a task carried out in the public interest".[199] In fact, the reach of the DA is more substantial than this,[200] among others, addressing also switching between data processing service providers,[201] interoperability,[202] and the *sui generis* database right.[203] Crucially, the DA would address also the use of "smart contracts" for data sharing, which is addressed in further detail below.[204]

[195]Defined as an information society services that comprises a mere conduit, caching or hosting service (Art. 3(g) DSA).
[196]Art. 9 DSA; illegal content is defined as "any information that, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with Union law or the law of any Member State which is in compliance with Union law, irrespective of the precise subject matter or nature of that law" (Art. 3(h) DSA), where the "the non-authorised use of copyright protected material" is an "illustrative example" thereof (Recital 12 DSA).
[197]Chapter III DSA; some of which are limited to "online platforms", which are defined as "a hosting service that, at the request of a recipient of the service, stores and disseminates information to the public, unless that activity is a minor and purely ancillary feature of another service or a minor functionality of the principal service and, for objective and technical reasons, cannot be used without that other service, and the integration of the feature or functionality into the other service is not a means to circumvent the applicability of this Regulation" (Art. 3(i) DSA).
[198]Consistent also with the DGA, the Data Act defines data as "any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audio-visual recording" (Art. 2(1) Data Act).
[199]Art. 1(1) DA.
[200]See regarding a comprehensive discussion: Josef Drexl and others, 'Position Statement of the Max Planck Institute for Innovation and Competition of 25 May 2022 on the Commission's Proposal of 23 February 2022 for a Regulation on Harmonised Rules on Fair Access to and Use of Data (Data Act)' (25 May 2022); see also Charlotte Ducuing et al (eds), 'CiTiP White Paper on the Data Act Proposal' (CiTiP Working Paper Series, 22 October 2022).
[201]Chapter VI DA.
[202]Chapter VIII DA.
[203]Chapter X DA.
[204]Art. 36 DA.

15 Annex VI: Copyright Legal Context

The reach of the EU copyright *acquis*, the component of the EU *acquis Communautaire* – the acquired body of EU regulatory material – that broadly addresses copyright and related rights, is vast. Copyright, in the taxonomy of the EU, is not to be confused with the demarcation between what are classed as the tradition of copyright and the tradition of the *droit d'auteur*.[205] Rather, the EU copyright *acquis* is developed based on key recognized instruments of international law, and the specific competences under which legislative acts of the EU have been adopted. The development of the EU copyright *acquis* has important implications for the overall contextualization of questions concerning potentially copyright- or related rights-protected material, which will be addressed below.

In regard to XR data and content, there are no specific rules regarding such technologies, or their usage contained within the copyright *acquis*. In light of this, an interpretation of the rules of the *acquis* in effect and their applicability to the specific identifiable fact pattern(s) distinguishable within XReco must be carried out. Finally, it should be clarified that the angles of legal analysis below are some of the main aspects that may fall within the scope of the copyright *acquis* and neither prejudices nor exhausts the breadth and depth of the discussion, considering both the dynamic of the technologies and the fact that there are technical steps involved that are not addressed in this section of the deliverable.

15.1 International Law and the EU Copyright Acquis

In the EU, copyright and related rights are protected under the national law of the Member States. The EU's copyright *acquis* therefore chiefly comprises legislation taking the form of Directives – that is, pieces of secondary legislation that is binding, as to the result to be achieved, upon each Member State to which it is addressed, leaving to the national authorities the choice of form and methods.[206] These Directives recognize the authority of international legal treaties in the domain of copyright and related rights, to which the EU copyright *acquis* contributes.[207] In its capacity as an international organization, the EU and its Member States are further required to comply with certain international treaties within the domain of copyright and related rights under Article 216 TFEU,[208] as well as the agreement of the European Economic Area Agreement.[209]

In this regard, the following treaties and provisions of international law are authoritative in the context of the EU copyright *acquis*:[210]

[205]Cf. Sam Ricketson and Jane Ginsburg, International Copyright and Neighbouring Rights: The Berne Convention and Beyond (2nd ed, OUP), 230-231.

[206]Art 288 Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C 326/47 (TFEU); there are also Regulations addressing copyright matters, suchas Regulation 2017/1563 [2017] OJ L 242/1 (Marrakech Regulation).

[207]E.g. Recital 19 Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L 167/10 (InfoSoc Directive).

[208]Article 216 TFEU states that: "(1.) The Union may conclude an agreement with one or more third countries or international organisations where the Treaties so provide or where the conclusion of an agreement is necessary in order to achieve, within the framework of the Union's policies, one of the objectives referred to in the Treaties, or is provided for in a legally binding Union act or is likely to affect common rules or alter their scope. (2.) Agreements concluded by the Union are binding upon the institutions of the Union and on its Member States."

[209]Art. 5 Protocol 28 Agreement on the European Economic Area - Final Act - Joint Declarations - Declarations by the Governments of the Member States of the Community and the EFTA States - Arrangements - Agreed Minutes - Declarations by one or several of the Contracting Parties of the Agreement on the European Economic Area [1994] OJ L 3/1 (EEA Agreement).

[210]These are not the only treaties, but are considered most relevant here.

- Berne Convention;[211]
- TRIPS Agreement;[212]
- WIPO Copyright Treaty.[213]

The adopted EU legislation in the field of copyright is understood here to comprise:

- Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Codified version) (Text with EEA relevance) (Software Directive)
- Directive 2006/115/EC of the European Parliament and of the Council of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property (codified version) (Rental and Lending Directive)
- Directive 93/83/EEC of 27 September 1993 on the coordination of certain rules concerning copyright and rights related to copyright applicable to satellite broadcasting and cable retransmission (Satellite and Cable Directive)
- Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights (codified version) (Term Directive)
- Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (Database Directive)
- Directive 2001/84/EC of the European Parliament and of the Council of 27 September 2001 on the resale right for the benefit of the author of an original work of art (Resale Right Directive)
- Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (InfoSoc Directive)
- Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works Text with EEA relevance (Orphan Works Directive)
- Directive(EU) 2017/1564 of the European Parliament and of the Council of 13 September 2017 on certain permitted uses of certain works and other subject matter protected by copyright and related rights for the benefit of persons who are blind, visually impaired or otherwise print-disabled and amending Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society (Marrakesh Directive)
- Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance) (CDSM Directive)

15.2 Principles of Copyright Law

15.2.1 Territoriality

Even in the EU single market, the rules of the copyright *acquis* have a territorial scope. A title in copyright exists in the country of origin,[214] that is, usually in the country where a work is first published,[215] though there are

[211]Berne Convention for the Protection of Literary and Artistic Works (Paris Act of 24 July 1971), as amended on 28 September 1979 (Berne Convention, BC).
[212] WTO Agreement on Trade-Related Aspects of Intellectual Property Rights (15 April 1994) 1869 UNTS 299, 33 ILM 1197 (TRIPS Agreement).
[213] WIPO Copyright Treaty (20 December 1996) 2186 UNTS 121, 36 ILM 65 (WCT).
[214]Art. 5 BC (9 TRIPS, 1 WCT)
[215]Art. 5(4)(a) BC.

derogating rules in the case of simultaneous publication and unpublished works.[216] Within the international framework, the national treatment requirement has been laid down for both copyright under the Berne Convention.[217] Within the EU, further discrimination is forbidden pursuant to Article 18 of the TFEU.[218] This means that the rights in a work originating in Member State A may not receive a less favourable protection than a work originating in Member State B.[219]

15.2.2 No Formalities

Under Art. 5(2) of the Berne Convention, the enjoyment and exercise of the rights to national treatment is required to not be made subject to any formality. This means that countries can still make certain formalities, such as registration of the work with a central body, beneficial for authors within their own country of origin of their work, but usually countries do not impose higher burdens on their own authors.[220]

15.2.3 Idea-expression Dichotomy

In copyright laws, a distinction is made between "ideas" and "expressions" .[221] As copyright protects works of authorship that are original expressions, whereas ideas themselves are generally not afforded protection.[222] In the context of the EU copyright *acquis*, this distinction is not present in all relevant legal instruments, yet references are made to the non-protectability of ideas and facts in the legislation and case law of the CJEU.[223]

15.2.4 Subsistence and Originality

The subject matter of the EU copyright *acquis* are works of authorship,[224] which, as the CJEU has clarified, need to be original in the sense of the author's own intellectual creation and be identified with sufficient precision and objectivity.[225]

Works are not defined by EU legislation but have been addressed by the case law of the CJEU. According to the CJEU, to qualify as a "work", the subject matter in question must fulfil the conditions of (1) an original subject matter (the author's own intellectual creation) must exist, and (2) the classification of "work" is reserved to the elements that are the expression of the author's own intellectual creation.[226] Statutorily, originality has been defined in the Software,[227] Database[228] and Term Directives[229] as encompassing an "author's own intellectual creation". The case law of the CJEU has expanded the ambit of originality to all subject matter covered by the

[216]Art. 5(4)(b) and (c) BC.
[217]Art. 5 Berne Convention, Artt. 4,5,6 Rome Convention; it should be noted that Malta is the sole EU Member State that is not a party to the Rome Convention. The EU and Malta are, however, parties to the WIPO Performances and Phonograms Treaty 2186 UNTS 121 (WPPT), which outlines the principle of national treatment in regard to performances and phonograms (Art. 4).
[218]Joined cases C-92/92 and C-326/92 *Phil Collins* [1993] ECLI:EU:C:1993:847.
[219]Case C-28/04 *Tod's* [2005] ECLI:EU:C:2005:418, para. 32.
[220]Ricketson and Ginsburg, 306.
[221]E.g. Art. 9(2) TRIPS Agreement.
[222]Unless protected by another regime, such as e.g. data protection or trade secrecy.
[223] For instance in the Software and DSM Directives; see Margoni and Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2022) 71 GRUR International 685, 689-690.
[224]Artt. 2 and 3 InfoSoc Directive.
[225]Case C-5/08 *Infopaq I* [2009] ECLI:EU:C:2009:465, para. 37; Case C-145/10 *Eva-Maria Painer*[2011] ECLI:EU:C:2011:798, para. 87; Case C-310/17 *Levola* [2018] ECLI:EU:C:2018:899, para. 40; Case C-683/17 *Cofemel* [2019] ECLI:EU:C:2019:721, para. 32; Case C-833/18 *Brompton Bicycle* [2020] ECLI:EU:C:2020:461, para. 25.
[226]*Cofemel*, para. 29.
[227] Art. 1 (3) and Recital 8 Software Directive.
[228]Artt. 3 (1) and 6 Database Directive.
[229] Recital 16 Term Directive.

InfoSoc Directive,[230] clarifying that the creative abilities of the author are expressed by making "free and creative choices".[231]

15.3 Exclusive (Economic) Rights

The rights that constitute copyright are traditionally divided into moral and economic rights.[232] The EU copyright *acquis* has been very active in the domain of economic rights but has not harmonised moral rights at all.[233] The scope of harmonised economic rights is different between copyright and related rights.

Under the EU copyright *acquis*, authors and/or the owners of copyright have exclusive economic rights in their works:

- The **right of reproduction**: the "exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part" for their works.[234] This means that by default, permission is required in order to make copies of the work of an author.
- The **right to communicate subject matter to the public**:[235] the exclusive right to transfer a work to an audience that is located elsewhere than where it is being transferred from. The right of communication to the public within the EU *acquis* does not cover "direct performances" such as in front of a live audience.[236] By default, in order to communicate a work to the public, permission of the author is required.
- The right of distribution: the "exclusive right to authorise or prohibit any form of distribution to the public by sale or otherwise" for their works.[237] The right of distribution is limited to the transfer of ownership of a copy,[238] meaning that the right of distribution is "exhausted" after the first transfer of ownership of a copy and does not apply subsequent transfers of that copy.[239] The distribution of works or copies of works requires the permission of the author.
- The Resale Right Directive harmonises the so-called "droit de suite": a royalty based on the sale price for any resale of their works.[240] This right only addresses original works of art that are graphic or plastic art made by the artist.[241]

15.3.1 Related rights

In addition to the exclusive rights of authors and/or copyright owners, there exist so-called "related rights". These can differ for the subject matter in question.

[230] *Infopaq I*, paras. 36-37.
[231] *Infopaq I*, paras. 36-37.
[232] Moral rights, not addressed here, are generally rights that an author of a work retains irrespective of any potential transfer of economic rights, such as the right of attribution or the right of integrity ("[independently] of the author's economic rights, and even after the transfer of the said rights, the author shall have the right to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honor or reputation." (Art. 6bis(1) BC).
[233] See for instance, Recital 9 InfoSoc Directive.
[234] Art. 2 InfoSoc Directive.
[235] Art. 3 (1) InfoSoc Directive; see also Art. 8 Rental and Lending Directive, Art. 1 SatCab Directive, Art. 5 Database Directive
[236] C-283/10 *Circul Globus Bucuresti* [2011] ECLI:EU:C:2011:772.
[237] Art. 4 InfoSoc Directive.
[238] Case C-456/06 Peek and Cloppenburg[2008] ECLI:EU:C:2008:232.
[239] Art. 5 (c) Database Directive; Art. 4 (2) Software Directive; Art. 4(2) InfoSoc Directive.
[240] Art. 1 Resale Right Directive.
[241] Art. 2 Resale Right Directive.

- Performing artists have exclusive rights to fixations of their performances.[242]
- Producers of phonograms have exclusive rights in their phonograms (sound recordings) .[243]
- Producers of films have exclusive rights in original and copies of their films.[244]
- Broadcasters have exclusive rights in fixations of their broadcasts. These include: the right of reproduction[245] and the right to make available to the public.[246] (see also Dir. 2006/115)
- Publishers of press publications have exclusive rights of reproduction and making available to the public for the online use of their press publications by information society service providers.[247]
- Makers of databases have rights in their databases "which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database" .[248] [sui generis right]

See also:2006/116/EC Protection of previously unpublished works (Art. 4) and Critical and scientific publications (Art. 5).

15.4 EU copyright contract law acquis: an overview

Licensing agreements, as private agreements, licenses are primarily governed by general contract law rules and are subject to the principle of contractual freedom. In addition, copyright licenses are frequently regulated by specific rules of copyright contract law contained in national copyright acts.

No EU harmonized legal framework exists for the general law of contracts (apart from consumer law). The EU copyright law *acquis* which is relevant for licensing contracts is mainly composed by the following blocks:

15.4.1 Principle of licensing

EU copyright law recognizes the **existence of copyright licensing contracts** as tools for copyright management and governs **the possibility** of rightsholders to grant licenses. This is established in a horizontal way for both copyright and related rights by the Directive 2001/29 (rec.30).[249] In particular, reproduction right, right of communication to the public of works and right of making available to the public other subject-matter for copyright and related rights rightsholders (authors, performers, phonogram and video producers and broadcasting organizations) and distribution right for authors may all be subject to the granting of licenses. This principle is inversely reflected in the requirement of licensing of right by different rightsholders of copyright and related rights as set out in rec. 2 of Directive 2014/26.[250]

[242] Art. 2(b) InfoSoc Directive
[243] Art. 2(c) InfoSoc Directive.
[244] Art. 2 (d) InfoSoc Directive.
[245] Art. 2 (e) InfoSoc Directive.
[246] By "sale or otherwise" (Art. 3(2)(d) InfoSoc Directive).
[247] Art. 15 DSM Directive.
[248] Art. 7 Database Directive.
[249] Recital 30 InfoSoc Directive: "The rights referred to in this Directive may be subject […] to the granting of contractual licenses, without prejudice to the relevant national legislation on copyright and related rights."
[250] Recital 2 Directive 2014/26: "The dissemination of content which is protected by copyright and related rights requires the licensing of rights by different holders of copyright and related rights, such as authors, performers, producers and publishers." See also *ibid.* Recital 37: "Directive 2001/29/EC requires that a license be obtained for each of the rights in the online exploitation of musical works."

The possibility of granting individual licenses is also recognized in many other EU copyright law instruments, with regard particular rights or types of use of works or other subject matter. More precisely:

Art. 3 para 1 of the Directive 93/83/EEC sets out that the "the authorization [for communication to the public by satellite of copyright works] may be acquired only by agreement." and Art. 8 para. 1 of the same Directive sets out that "retransmission takes place on the basis of individual or collective contractual agreements between copyright owners, holders of related rights and cable operators."

Article 7 para. 3 of the Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases sets out that the *sui generis right* of a database maker may be transferred, assigned or granted **under contractual license.**

Article 3 para. 3 of the Directive 2006/115/EC sets out that **rental and lending rights** of works or other subject-matter (author's rights, in respect of the original and copies of his work, performer's rights in respect of fixations of his performance, phonogram producer's rights, in respect of his phonograms, producer's of the first fixation of a film in respect of the original and copies of his film) may be subject to the granting of contractual licenses.

Art. 9 para. 4 of Directive 2006/115/EC sets out that the **distribution right**, recognized to performers, phonogram and film producers and broadcasting organizations, may be transferred, assigned or subject to the granting of contractual licenses.

Art. 5 para. 2 of the Directive 2019/790 [CDSMD] makes implicit reference to the granting of **educational licenses**, i.e. licenses authorizing acts of digital use of works and other subject matter for the sole purpose of illustration for teaching, to the extent justified by the non-commercial purpose to be achieved.

Article 17 para. 1 CDSMD makes reference to the conclusion of a licensing agreement between authors and related rights rightsholders and online content-sharing service provider, in order to communicate to the public or make available to the public works or other subject matter. According to Article 17 para. 2 CDSM, Member States shall provide that, where an online content-sharing service provider obtains an authorisation, for instance by concluding a licensing agreement, that authorisation shall also cover acts carried out **by users** of the services falling within the scope of Article 3 of Directive 2001/29/EC when they are not acting on a commercial basis or where their activity does not generate significant revenues.

Reference to the licensing principle is also made within Recitals of EU Copyright Law instruments, such as

Rec. 18 of the Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, which makes reference to the possibility for author or the holder of a related right to permit some of his works or subject matter to be included in a database pursuant to a non-exclusive agreement.

Rec. 11 of the Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works, which makes reference to the licensing agreement for use of cinematographic and audiovisual works and phonograms contained in the archives of public-service broadcasting organizations by these organizations.

The principle of granting licenses tacitly reflects the premise of contractual freedom. However, explicit reference in the **principle of contractual freedom** in the licensing context is rarely made within the EU Copyright acquis.[251]

On another level, the CJEU (Case C-533/07 Falco Privatstiftung and Thomas Rabitsch v Gisela Weller-Lindhorsruled) ruled that the second indent of Article 5(1)(b) of Regulation No 44/2001 must be interpreted as meaning that a contract under which the owner of an intellectual property right (copyright licensor) grants its contractual partner the right to use that right in return for remuneration (a license agreement) is not a contract for the provision of services within the meaning of that provision. According to the court, the obligation of exploitation is not an essential feature of such contract (para. 44). On the contrary, the only obligation which the owner of the right granted (the licensor) undertakes with regard to its contractual partner is **not to challenge** the use of that right by the latter: the owner of an intellectual property right (who receives the remuneration-license fee) does not perform any service in granting a right to use that property and undertakes merely to permit the licensee to exploit that right freely (para. 31).

15.4.2 Rules on copyright and related rights licensing negotiations

EU copyright law acquis lays down rules related with the negotiation of individual licensing contracts in the following provisions:

According to article 12 para. 1 of the Directive 93/83/EEC **"**[T]he parties enter and conduct negotiations regarding authorization for cable retransmission in good faith and do not prevent or hinder negotiation without valid justification."

According to Art. 13 CDSMD "Member States shall ensure that parties facing difficulties related to the licensing of rights when seeking to conclude an agreement for the purpose of making available audiovisual works on video-on-demand services may rely on the assistance of an impartial body or of mediators."

Block of protective Rules for licenses granted by authors and performers

A relatively extensive set of substantial rules regulating copyright licensing contracts has been recently introduced by Art. 18-23 CDSMD [CHAPTER 3 Fair remuneration in exploitation contracts of authors and performers]. These rules apply only in onerous exploitation contracts, including licenses for the exploitation of works or performances, which are signed, at a primary level, **between the author or the performer** (natural persons and their own company[252]) and a third-party exploiter. The rationale behind this limitation of the scope of these licensing rules resides in the consideration of authors and performers as the weak party of such exploitation contracts (including licencing contracts intending to the exploitation of works/performances)[253] and need of protection that subsequently derives therefrom.

[251] See esArt. 18 para. 2 CDSMD: "In the implementation in national law of the principle set out in paragraph 1, Member States shall be free to use different mechanisms and take into account the **principle of contractual freedom** and a fair balance of rights and interests." and

Reg. 2017/1128 rec. (29): "A holder of copyright, related rights, or any other rights in the content of an online content service should remain able to exercise contractual freedom to authorise such content to be provided, accessed and used under this Regulation without verification of the Member State of residence."

[252] Rec. 72 CDSDM.

[253] Rec. 72 CDSMD.

That need for protection does not arise where the contractual counterpart acts as an end user and does not exploit the work or performance itself.[254] In addition, the need of information in assessing the economic value of rights does not arise where the exploitation has ceased, or where the author or performer has granted a license to the general public without remuneration.[255] Last, Articles 18 to 22 CDSMD do not apply to authors of a computer program within the meaning of Article 2 of Directive 2009/24/EC (Article 23 para. 2 CDSMD).

This set of rules is comprised by the below components:

a. Principle of appropriate and proportionate remuneration

Article 18 para. 1 CDSMD sets out that:

"Member States shall ensure that where authors and performers license or transfer their exclusive rights for the exploitation of their works or other subject matter, they are entitled to receive **appropriate and proportionate remuneration**".[256]

The assessment of the appropriate and proportionate character of the remuneration is made with regards the actual or potential economic value of the licensed or transferred rights, taking into account the author's or **performer's contribution to the overall work** or other subject matter and all other circumstances of the case, such as **market practices** or the actual exploitation of the work.[257] A lump sum payment can also exceptionally constitute proportionate remuneration.[258] Moreover, the principle of appropriate and proportionate remuneration does not preclude the granting of licenses without consideration (**free licenses**),[259] such as the granting of Creative Commons licenses. But in that case, these licenses would not be considered as "licenses for the exploitation of works" in return for remuneration, where the principle of appropriate and proportionate remuneration would apply.

b. Transparency obligation of licensees

Article 19 para. 1 CDSMD sets out a minimum[260] transparency obligation for licensees in the framework of individual licenses,[261] according to which:

"Member States shall ensure that authors and performers receive on a regular basis, at least once a year, and taking into account the specificities of each sector, up to date, relevant and comprehensive information on the exploitation of their works and performances from **the parties to whom they have licensed** or transferred their rights, or their successors in title, in particular as regards modes of exploitation, **all revenues generated** and remuneration due."

Moreover, in case that licensed rights are further sublicensed, authors and performers or their representatives shall, at their request, receive **from sub-licensees additional information**, in the event that **their first contractual**

[254] Rec. 72 CDSMD.
[255] Rec. 74 CDSMD.
[256] This principle was initially anticipated in InfoSoc Directive, see Recital 10: [A]uthors have to receive an appropriate reward for the use of their work" and Recital n° 45: "The exceptions and limitations referred to in Article 5(2), (3) and (4) should not, however, prevent the definition of contractual relations designed to ensure fair compensation for the rightsholders insofar as permitted by national law."
[257] Rec. 73 CDSMD.
[258] Rec. 73 CDSMD.
[259] Rec. 82 CDSMD.
[260] Rec. 76 CDSMD: "Member States should have the option, in compliance with Union law, to provide for further measures to ensure transparency for authors and performers." and Rec. 77 CDSMD: "Such agreements should ensure that authors and performers have the same level of transparency as or a higher level of transparency than the **minimum requirements** provided for in this Directive." This means that the rules in question should be seen as rules of full harmonisation.
[261] Rec. 77 CDSMD, *in fine*.

counterpart does not hold all the information that would be necessary for providing the above information and high level of transparency. Where that additional information is requested, the first contractual counterpart of authors and performers shall provide information on the identity of those sub-licensees (Art. 19 para. 2 CDSMD). This exploitation-related information is duly provided to authors and performers by sub-licensees who exploit the rights (Rec. 76 CDSMD[262]).

The above transparency obligation is reflected as right to request information to the benefit of authors and performers.[263] This information is deemed necessary for authors and rightsholders in order to be able to assess **the economic value of their rights**, compared to the remuneration received,[264] esp. in comparison with relevant revenues for the licensee, derived from the exploitation of licensed content.

The information should refer to recent data and in a way that it covers all sources of revenues relevant to the case, on all modes of ongoing exploitation and on all relevant revenues worldwide, including, where applicable, merchandising revenues. Information shall be provided regularly, according to the trends of the relevant sector, but at least annually. The information should be provided in a manner that is comprehensible to the author or performer and it should allow the effective assessment of the economic value of the rights in question.[265]

Obligation of transparency that is imposed upon licensees and sub-licensees is subject to some limitations:

Member States may provide that any request to sub-licensees pursuant to the first subparagraph is made directly or indirectly through the contractual counterpart of the author or the performer (Art. 19 para. 2 CDSMD).

Moreover, transparency obligation shall **be proportionate and effective** in ensuring a **high level of transparency** in every sector. Member States may provide that in duly justified cases where the administrative burden resulting from the obligation set out in paragraph 1 would become disproportionate in the light of the revenues generated by the exploitation of the work or performance, the obligation is limited to the types and level of information that **can reasonably be expected in such cases**. (Art. 19 para. 3 CDSMD). This means that Member States shall take into account the specificities of different content sectors and that all relevant stakeholders should be involved when deciding on such sector-specific obligations (rec. 77 CDSMD).

Also, Member States may decide that the obligation of transparency does not apply when the contribution of the author or performer **is not significant having regard to the overall work or performance**, unless the author or performer demonstrates that he or she requires the information for the exercise of his or her rights under Article 20(1) and requests the information for that purpose (Art. 19 para. 3 CDSMD).

A specific transparency obligation is also set out by Article 17 para. 8 CDSMD according to which Member States shall provide that online content-sharing service providers provide rightsholders, at their request, with **adequate information** on the functioning of their practices with regard to the cooperation referred to in paragraph 4 and, where licensing agreements are concluded between service providers and rightsholders, **information on the use**

[262] Rec. 76 CDSMD: "In order to ensure that exploitation-related information is duly provided to authors and performers also in cases where the rights have been sub-licensed to other parties who exploit the rights [...]".
[263] Rec. 76 CDSMD: Authors and performers, and their contractual counterparts, should be able to agree to keep the shared information confidential, but authors and performers should always be able to use the shared information for the purpose of exercising ***their rights*** under this Directive
[264] Rec. 74 and 75 CDSMD: (75) As authors and performers tend to be in the weaker contractual position when they grant licenses or transfer their rights, they need information to assess the continued economic value of their rights, compared to the remuneration received for their license or transfer, but they often face a lack of transparency. Therefore, the sharing of adequate and accurate information by their contractual counterparts or their successors in title is important for the transparency and balance in the system governing the remuneration of authors and performers.
[265] Rec. 75 CDSMD.

of content covered by the agreements. This obligation is applied in the context of a licensing agreement signed by any rightsholder (not only author or performer).

c. Remuneration adjustment mechanism ("best-seller" provision)

Article 20 CDSMD sets out a contract adjustment mechanism for the event that the economic value of the rights turns out to be significantly higher than initially estimated when authors or performers signed and exploitation contract in return for remuneration.[266] According to Article 20 CDSMD Member States shall ensure that, in the absence of an applicable collective bargaining agreement providing for a mechanism comparable to that set out in this Article, authors and performers or their representatives are entitled **to claim additional, appropriate and fair remuneration** from the party with whom they entered into a contract for the exploitation of their rights, or from the successors in title of such party, when the remuneration originally agreed turns out to be **disproportionately low** compared to **all the subsequent relevant revenues derived from the exploitation** of the works or performances.

The assessment of the situation should take account of the specific circumstances of each case, **including the contribution of the author or performer**, as well as of the specificities and remuneration practices in the different content sectors, and whether the contract is based on a collective bargaining agreement.

All revenues relevant to the case in question, including, where applicable, merchandising revenues, should be taken into account for the assessment of whether the remuneration is disproportionately low.[267]

Paragraph 1 of this Article applies in the context of license for the exploitation of licensed content especially when the license is granted for a long duration[268] (such as for the whole term of protection). but shall not apply to agreements concluded by CMOs and IMEs or by other entities that are already subject to the national rules implementing that Directive.

Where the parties of the license do not agree on the adjustment of the remuneration, the author or performer should be entitled to bring a claim before a court or other competent authority.[269]

d. Mechanism for the revocation of rights

Article 22 CDSMD provides for a right of revocation benefitting authors or performers that licensed on an exclusive basis right according to which the author or performer may revoke in whole or in part the license where there **is a lack of exploitation of that work or other protected subject matter.**

The revocation will allow the author or performer to transfer or license their rights to another person.[270]

The right of revocation may be exercised by the licensor (author or performer) in accordance with certain procedural requirements (notification, deadline) and only after a certain (reasonable[271]) period of time following the conclusion of the license or of the transfer agreement (Art. 22 para. 3 CDSMD; Rec. 80 CDSMD). After the expiry of the deadline set, the author or performer may choose **to terminate the exclusivity of the contract instead of revoking the license or of the rights** (Art. 22 para. 2 CDSMD).

[266] Rec. 78 CDSMD.
[267] Rec. 78 CDSMD.
[268] Rec. 78 CDSMD.
[269] Rec. 78 CDSMD.
[270] Rec. 80 CDSMD.
[271] Rec. 80 CDSMD.

The revocation right shall not apply if the lack of exploitation is predominantly due to circumstances that the author or the performer can reasonably be expected to remedy (Art. 22 para. 4 CDSMD). Member States may exclude works or other subject matter from the application of the revocation mechanism if such works or other subject matter usually contain contributions of a plurality of authors or performers (Art. 22 para. 2 CDSMD). In addition, Member States may provide that the revocation mechanism can only apply within a specific time frame, where such restriction is duly justified by the specificities of the sector or of the type of work or other subject matter concerned (Art. 22 para. 2 CDSMD).

Last, specific provisions for the regulation of the revocation mechanism may be provided for in national law, per creative sector or per the type of work or performance, in particular providing for time frames for the right of revocation. Specific regulation may be enacted also in case of a work containing the contribution of more than one author or performer, taking into account the relative **importance of the individual contributions** and the legitimate interests of all authors and performers affected by the application of the revocation mechanism by an individual author or performer (Art. 22 para. 2 CDSMD; rec. 80 CDSMD).

Member States may provide that any contractual provision derogating from the revocation mechanism is enforceable only if it is based on a collective bargaining agreement (Art. 22 par. 5; rec. 80 CDSMD).

Rules governing collective licenses

Directive 2014/26 recognizes the possibility of granting a license as way for establishing a direct legal relationship with a Collective management organization (Art. 3 a) Dir. 2014/26[272] and Art. 7 Dir. 2014/26) or an independent management entity (Art. 3 b) Dir. 2014/26[273]), for the purpose of mandating the collective management of their rights. This type of license is not relevant in the XReco context, to the extent that XReco platform is not intended to function as a CMO or IME, and the relevant rules provided for by the Directive 2014/26 will not be tackled.

Directive 2014/26 provides for a set of rules regulating the licenses that are granted by CMOs to users that cover collectively the represented repertoire by the Licensor-CMO (**collective licenses**). This set of rules is composed by the following blocks:

a. Principle of good faith and fairness in negotiation and determination of tariffs of collective licenses

Article 16 para. 1 Dir. 2014/26 introduces the rule according to which Member States shall ensure that collective management organisations and users conduct negotiations for the licensing of rights in good faith. Collective management organisations and users shall provide each other with all necessary information. Within that context, Collective management organisations shall reply without undue delay to requests from users, (including by electronic means, Art. 16 para. 4 Dir. 2014/26), indicating, inter alia, the information needed in order for the collective management organisation to offer a license (Art. 16 para. 3 Dir. 2014/26).

In addition, according to Art. 16 para. 2 Dir. 2014/26, licensing terms shall be based on objective and non-discriminatory criteria. In particular, tariffs applied by CMOs should be fair, in order to ensure that users can obtain licenses and should be determined on the basis of objective and non-discriminatory criteria and be reasonable in relation to, inter alia, the economic value of the use of the rights in trade, taking into account **the**

[272] Art. 3 a) Dir. 2014/26: 'collective management organisation' means any organisation which is authorised by law or by way of assignment, license or any other contractual arrangement to manage copyright or rights related to copyright on behalf of more than one rightsholder, for the collective benefit of those rightsholders [...]".

[273] Art. 3 (b) Dir. 2014/26: 'independent management entity' means any organisation which is authorised by law or by way of assignment, license or any other contractual arrangement to manage copyright or rights related to copyright on behalf of more than one rightsholder, for the collective benefit of those rightsholders,

nature and scope of the use of the work and other subject-matter, as well as in relation to the economic value of the service provided by the collective management organization, in order to ensure appropriate remuneration of rightsholders. Collective management organisations shall inform the user concerned of the criteria used for the setting of those tariffs.

Upon receipt of all relevant information by a user, the collective management organisation shall, without undue delay, either offer a license or provide the user with a reasoned statement explaining why it does not intend to license a particular service (Art. 16 para. 3 Dir. 2014/26).

Collective management organisations shall not be required to use, as a precedent for other online services, **individualized licensing terms** agreed with a user where the user is providing a new type of online service (innovative online services) which has been available to the public in the Union for less than three years. Dir. 2014/26 (Art. 16 para. 2 Dir. 2014/26; Rec. 32 Dir. 2014/26).

b. Reporting obligation of the licensee (user)

According to Article 17 Dir. 2014/26, Member States shall adopt provisions to ensure that users provide a collective management organisation, within an agreed or pre-established time and in an agreed or pre-established format, with such relevant information at their disposal on the use of the rights represented by the collective management organisation as is necessary for the collection of rights revenue and for the distribution and payment of amounts due to rightsholders.

When deciding on the format for the provision of such information, collective management organisations and users shall take into account, as far as possible, voluntary industry standards.

A collective management organisation shall allow users to communicate with it by electronic means, including, where appropriate, for the purpose of reporting on the use of the license (Art. 16 para. 4 Dir. 2014/26).

c. Non commercial individual licenses

Article 5 para. 3 Dir. 2014/26 offers the possibility for rightsholders that have mandated the representation of their rights/repertoire by a CMO to grant (individual) licenses for non-commercial uses of any rights, categories of rights or types of works and other subject-matter that they may choose, even though their rights are still managed collectively by the CMO.

d. Rules on multi-territorial licensing of online music rights.

Articles 23-32 Directive 2014/26 provide specific rules regulating the granting and content of multiterritorial licenses of online music rights by CMOs. These rules might be relevant in the XReco context, insofar musical work/content is at stake.

e. Extended collective licensing (ECL)

i) General ECL framework

Article 12 CDSMD complements the rules of Dir. 2014/26 by introducing the collective licensing with an extended effect on a territorial basis. By means of this provision, Member States **may** provide that where a collective management organisation enters into a licensing agreement for the exploitation of works or other subject matter, such an agreement can be extended to apply to the rights of rightsholders who have not authorised that collective management organisation to represent them ('outsiders'), or with respect to such an agreement, the organisation has a legal mandate or is presumed to represent rightsholders who have not authorised the

organisation accordingly, when certain safeguards are fulfilled (sufficiently representative CMO, non discrimination, opt-out possibility, well-defined areas of use covered by voluntary collective management of rights, where obtaining authorisations from rightsholders on an individual basis is typically onerous and impractical to a degree that makes the required licensing transaction unlikely) (Art. 12 para. 2 and 3 CDSMD).

ii) Specific ECL framework

Article 12 CDSMD does not affect the application of collective licensing mechanisms with an extended effect in accordance with other provisions of Union law, which include the non mandatory ECL licensing mechanism that a Member State may provide for the agreement between a collecting society and a broadcasting organization for the communication to the public of works by satellite (Art. 3 para. 2 Directive 93/83/EEC), the mandatory ECL mechanism for the exercise of the cable retransmission right of copyright owners and holders of related rights (Art. 9 Directive 93/83/EEC), or the licensing mechanism for licenses with extended effect covering cross-border (pan-European) uses of out-of-commerce works and signed between representative CMOs and cultural heritage institutes (Article 8 and 9 CDSMD).

Mandatory law

In certain cases, EU law provides explicitly for the mandatory nature of some rules regulating licensing contracts, meaning that, these rules will constitute *ius cogens* also in the framework of their transposition in national law, and parties will not be able to derogate by means of contractual agreement.[274]

This is the case for transparency obligation (Art. 19 CDSMD) and Remuneration adjustment mechanism (Art. 20 CDSMD). According to Art. 23 para. 1 CDSMD, Member States shall ensure that any contractual provision that prevents compliance with Articles 19, 20 and 21 **shall be unenforceable** in relation to authors and performers. The mandatory nature of these rules refers both to licenses between authors, performers and their contractual counterparts [licensees], or in agreements between those counterparts [licensees] and third parties.[275]

Similarly, Article 7 of the Reg. 2017/1128 provides that any contractual provisions, including those between providers of online content services and holders of copyright or related rights or those holding any other rights in the content of online content services, as well as those between such providers and their subscribers, which are contrary to this Regulation, including those which prohibit cross-border portability of online content services or limit such portability to a specific time period, shall be unenforceable.

In other cases, the content of a licensing agreement is indirectly determined by the mandatory nature of certain exceptions and limitations under EU law. Accordingly, any contractual provision contrary to the exception of text and data mining for the purposes of scientific research (Art. 3 CDSMD), to the exception related with the Use of works and other subject matter in digital and cross-border teaching activities (Art. 5 CDSMD) or the exception related with preservation of cultural heritage (Art. 6 CDSMD) shall be unenforceable (Art. 7 CDSMD).[276]

[274] Rec. 81 CDSMD.

[275] Rec. 81 CDSMD.

[276] Similarly, in the case of computer program licensing (not relevant for XReco), any contractual provisions contrary to Article 6 [decomplilation] or to the exceptions provided for in Article 5(2) [back-up copy] and (3) ["study acts"] shall be null and void, insofar, inter alia, those acts are performed by the licensee or by another person having a right to use a copy of a program, or on their behalf by a person authorised to do so (Art. 8 of the Directive 2009/24/EC).

Application of technological measures of protection

Dir. 2001/29 recognizes the possibility for rightsholders to make use of technological measures of protection, i.e. the application of any technology, device or component that, in the normal course of its operation, is designed to prevent or restrict acts, in respect of works or other subject-matter, which are not authorised by the rightsholder of any copyright or any right related to copyright (Art. 6 para. 3 Dir. 2001/29). This of course may take place by means of a licensing agreement. In that regard, technological measures will safeguard the respect of the scope of the license by the licensee.

Technological measures shall enjoy adequate legal protection against the circumvention by Member State's legislation (Art. 6 para. 1 Dir. 2001/29). Notwithstanding the legal protection, in the absence of voluntary measures taken by rightsholders, including agreements between rightsholders and other parties concerned, Member States shall take appropriate measures to ensure that rightsholders make available to the beneficiary of an exception or limitation provided for in national law in accordance with Article 5(2)(a), (2)(c), (2)(d), (2)(e), (3)(a), (3)(b) or (3)(e) the means of benefiting from that exception or limitation, to the extent necessary to benefit from that exception or limitation and where that beneficiary has legal access to the protected work or subject-matter concerned. The provisions of the first and second subparagraphs shall not apply to works or other subject-matter made available to the public on agreed contractual terms in such a way that members of the public may access them from a place and at a time individually chosen by them.

The first, third and fifth subparagraphs of Article 6(4) of Directive 2001/29/EC shall apply to TDM exceptions of Art. 3 and 4 of the CDSMD as well as the exception related with use for teaching and the exception related with preservation of cultural heritage (Art. 7 para. 2 CDSMD).

Locus standi - Right to sue

According to Art. 4 (b) of Directive 2004/48 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 29 April 2004 on the enforcement of intellectual property rights, licensees are e entitled to apply for the application of the measures, procedures and remedies, in so far as permitted by and in accordance with the provisions of the applicable law.

16 Annex VII: Licensing and Rights Management

16.1 The concept of rights management: a general approach

Rights management refers to copyright and related rights management. Generally speaking, the concept **of rights management** has a different meaning depending on the viewpoint adopted:

- **Rightsholders' viewpoint**: In case of **management of rights** by the persons who **own or acquired these rights (rightsholders)**, then rights management is connected with the **definition** (imposition) of terms of access and use of own content which is protected by copyright and related rights (hereinafter: protected content). Under this approach, the term "management of rights" refers to the means by which copyright and related rights are administered, i.e., **licensed, assigned or remunerated** for any type of use,[277] including **licensing policy/strategy**. In addition, rights management and licensing strategy includes auditing of users and monitoring of the use of rights, as well as enforcement of rights.[278] According to Art. 7 para. 2 Dir. 2001/29, the expression "rights-management information" means any information provided by rightsholders which identifies the work or other subject-matter, the author or any other rightsholder, or **information about the terms and conditions of use of the work or other subject-matter**, and any numbers or codes that represent such information.
- **User's viewpoint:** In case of **management of third-party rights**, then rights management is connected with the **respect** of the terms of access and use of protected content as set by the rightsholder(s) thereof and/or the law by the person who desires to undertake this access/use or exploitation **(the user)**. Under this approach, rights management refers also to **rights clearance** and to respect of **lawful access** to protected content by users.

Both aspects of rights management, as described above, are two faces of the same coin: they stem from the absolute and exclusive character of copyright and related rights as recognized under EU law, as well as from the principle of safeguarding high level of protection of copyright and related rights in the EU.[279] In fact, (primary, i.e. initial or secondary, i.e. transferee or licensee) rightsholder of copyright or related rights has the power:

(a) to discretionally define the terms of access and use of own content that falls within the scope of own economic rights (licensing strategy definition), including monetization terms for the authorized access and use;

(b) to enforce her rights in case that access and use of protected content is made without authorisation or in deviation of the terms of access and use set by the rightsholder

Definition of terms of access and use is only relevant in case that **copyright-sensitive acts** (acts of use or exploitation) are at stake.

Accordingly, rights management, as expression of an administration rights policy or form of rights clearance, is primarily made on a voluntary basis, on the grounds of **voluntary agreements** (such as licensing agreements, see below, under Section **Fehler! Verweisquelle konnte nicht gefunden werden.** between rightsholders and users. Alternatively, rights management (administration of rights/rights clearance) may be defined by statutory rules,

[277] Communication from the Commission to the Council, the European Parliament and the European Economic and Social - Committee The Management of Copyright and Related Rights in the Internal Market (Text with EEA relevance) /* COM/2004/0261 final */, under "Executive Summary".
[278] Cf. Dir. 2014/26, rec. 2.
[279] Infosoc Directive, Rec. 4.

when access and use of protected content falls within the scope of **exceptions and limitations, including mandatory licensing** (statutory licenses) of copyright and/or related rights. In this case, users will correspond to beneficiaries of exceptions/limitations.

Given the nature and restrictive interpretation of exceptions and limitations, licensing may be seen as the default solution for rights management, whereas rights management based on exceptions and limitations is an alternative.

On another level, rights may be managed (i.e., administered) individually (**individual management**) by their initial or secondary owner/rightsholder, or collectively, by a competent collective management organisation (**collective management**). Individual rights management refers to the targeted management of rights (administration of rights/rights clearance) that belong to individually identified rightsholders.[280] In case of collective management, all represented rights from various rightsholders are altogether managed by the CMO on a collective basis without discriminations. It is normally for the rightsholder to choose between the individual or collective management of his rights, unless Member States provide otherwise, in compliance with Union law and the international obligations of the Union and its Member States.[281]

Under copyright law, the nature of the exclusive rights outlined above is such that any person that is not the relevant rightsholder regarding a certain work or other subject matter requires an authorisation to perform certain acts vis-à-vis the work or other subject matter in question. These authorisations can be provided by statute (exceptions and limitations to copyright), or by the rightsholder themselves (usually in the form of a license).

16.2 Voluntary licensing

16.2.1 General licensing principles and main building blocks

The concept of license/licensing agreement

Licensing refers to the granting of licenses. A license is a contract/agreement (licensing agreement) involving two parties: the holder of rights over protected content (=**licensor**) and the acquirer of rights (=**licensee**). Moreover, any license is attached to specific protected content (**licensed content**). As a minimum, a license provides the permission (authorisation[282]) of licensor to licensee to perform an copyright-sensitive **act of use or exploitation[283]** associated with protected content, which without that permission would be an infringement of copyright or a related right, because this act is covered by the content of exclusive copyright/related rights[284] (**act protected by copyright, by related rights or by the sui generis database right or by all of them)**. Accordingly, from a rights management perspective, the licensor coincides with the rightsholder and licensee to the user of the licensed content (see above under 1).

[280] Cf. Kaya Köklü, Individual Licensing of Copyrighted Works, In: R.M. Hilty and K.-C. Liu (eds.), Remuneration of Copyright Owners, Berlin/Heidelberg (Springer 2017), p177 ff.

[281] Directive 2014/26, rec. 2.

[282] From an EU Law point of view, a licensing agreement may be seen as on possible form to provide such authorisation, see e.g. CDSMD Rec. 64: "online content-sharing service providers should obtain an authorisation, including via a licensing agreement, from the relevant rightsholders."

[283] The dissociation between exploitation and use of protected (licensed) content is based on the type of intended use of the content (i.e. passive end use or further onerous dissemination). Accordingly, licensee would be seen as a commercial or **end user** of the licensed content and the license granted as **commercial license** or **end-user license**. See in that regard CDSMD, rec. 72: "[...] the contractual counterpart acts as an end user and does not exploit the work [...]".

[284] Communication from the Commission to the Council, the European Parliament and the European Economic and Social - Committee The Management of Copyright and Related Rights in the Internal Market (Text with EEA relevance) /* COM/2004/0261 final, ftnote 12.

It is possible that a single protected content is co-authored or owned by **multiple rightsholders** who own a share of rights over it. In such cases, a permission stemming from all involved rightsholders/Licensors would be necessary[285] (e.g., by means of **joint licenses[286]).**

Accordingly, a license is a contract establishing a legal relationship between licensor and licensee, by which the licensor exercises its own or acquired exclusive rights to permit (authorise) the licensee to make use (=typically passive end use) / or exploit (=typically lucrative use) the licensed content, according to the special terms of use/exploitation that are agreed between the parties.[287]

The agreement over the content of the license may be subject to negotiations between the parties.[288] It may also be offered as a standard contract by the Licensor, such as in case of wrap-up license agreements offered and signed online (e.g., by means of acceptance of terms and conditions by means of a click).

(Voluntary/contractual) licensing of rights is the most widespread means of rights management.[289] Licensing is typically distinguished from assignment (i.e. transfer) of rights in an exclusive and definitive manner.[290] Voluntary licensing is also distinguished from mandatory licensing i.e. statutory licensing against remuneration/compensation, which corresponds to a limitation of rightsholder's exclusive power to allow and forbid acts of use or exploitation.[291]

Parameterization of a license: the main building blocks

The parameterization of a license refers to the definition of several **building blocks** that usually compose the content of a licensing agreement.[292] The basic building blocks of a license, reflecting the content of the contractual arrangement, are the following:

The presence and value of the below building blocks determines the attributes of each licensing agreement.

a. Exclusivity

A license and, accordingly, use/exploitation of the licensed content by the licensee may be exclusive or non-exclusive (simple).[293] In case of exclusive license, the licensee becomes exclusive user/exploiter of the licensed content (within the scope of the license), excluding also the licensor from performing concurring use/exploitation

[285] In the existence of multiple rightsholders over a single item [and a fortiori, in case of multiple items], obtaining licenses on an individual basis may be onerous and impractical, cf. Art. 12 para. 2 CDSMD.
[286] Cf. CDSMD, rec. 33, in the context of collective licensing for out-of-commerce works.
[287] A. Strowel and B. Vanbrabant, Copyright licensing: a European view in: J. de Werra (ed.), Research Handbook on IP Licensing, (Edward Elgar Editors 2013), p29-53 and es34.
[288] Negotiations may of course fail. *Cf*. CDSMD, rec. 32: "A lack of agreement on the conditions of the license should not be interpreted as a lack of availability of licensing solutions."
[289] Communication from the Commission to the Council, the European Parliament and the European Economic and Social - Committee The Management of Copyright and Related Rights in the Internal Market (Text with EEA relevance) /* COM/2004/0261 final, under 1.1.1. *Of course,* in some industries, such as books publishing, assignment is more common, see A. StrowelandB. Vanbrabant, op.cit., 34.
[290] Communication from the Commission to the Council, the European Parliament and the European Economic and Social - Committee The Management of Copyright and Related Rights in the Internal Market (Text with EEA relevance) /* COM/2004/0261 final, ftnote 12; A. Strowel and B. Vanbrabant, op.cit,. 34.
[291] See among others Ch. Geiger, Statutory Licenses as Enabler of Creative Uses (December 19, 2015). In: R.M. Hilty and K.-C. Liu (eds.), Remuneration of Copyright Owners, Berlin/Heidelberg, Springer, 2017, p305-327, Max Planck Institute for Innovation and Competition Research Paper No. 15-14, Available at SSRN: https://ssrn.com/abstract=2701862, 4 ff.
[292] Cf. A. StrowelandB. Vanbrabant, ocit., 30.
[293] Communication from the Commission to the Council, the European Parliament and the European Economic and Social - Committee The Management of Copyright and Related Rights in the Internal Market (Text with EEA relevance) /* COM/2004/0261 final, under 1.1.1.; Reg. 2017/1128, rec.10: "The acquisition of a license for relevant rights is not always possible, in particular when rights in content are licensed on an exclusive basis."

of the licensed material. In case that licensor is also allowed to use/exploit the licensed content in parallel with the licensee, exclusivity refers to the quality of licensee (sole license).[294]

b. Limitations in scope

Licensing (and, accordingly, use/exploitation of the copyrighted content) may be subject to limitations. More precisely, any license, either exclusive or nonexclusive may be parameterized with regards its scope. The less the limitations, the broader the scope of licensed rights. Accordingly, protected content may be licensed **in whole or in pArt.** The case of an unlimited exclusive license hardly differentiates from a functional point of view from an assignment.[295]

Beyond the limitation regarding the object of a license, i.e., the definition/specification of the individual content (work or other protected subject matter) covered by the license, the license may be also limited in scope. Restrictions mainly refer to:

(1) Types of use covered by the license [exploitation modes] (e.g., reproduction, modification, derivative reuse, communication to the public etc.).[296] The specification of modes of exploitation may be more detailed (e.g. translation only in one language)[297] and may vary, depending on the type of licensed content.[298]

(2) Territory [geographical scope] (country/-ies)[299] (territorial licensing[300]). Several territories may be covered (multiterritorial license).[301]

(3) Term (Duration).[302]

(4) Purpose of use of licensed content (fields of use[303] or nature of use[304]) (e.g., commercial/non commercial,[305] research etc)

(5) Means of use of licensed content (e.g., digital, analog etc.)

[294] Cf. CDSM rec. 80: "Where those rights have been transferred on an exclusive basis, authors and performers cannot turn to another partner to exploit their works or performances."
[295] See e.g. A. StrowelandB. Vanbrabant, ocit., 34.
[296] Cf. CDSMD rec. 40 in the context of licensing out of commerce works: "Contracting cultural heritage institutions and collective management organisations should remain free to agree on the territorial scope of licenses, including the option of covering all Member States, the license fee and the uses allowed.
[297] Cf. CDSMD rec. 10: "[…]the terms of the licenses could exclude text and data mining." and CDSMD rec. 23: "[…] licenses for material that is primarily intended for the educational market […]."; placement of hyperlinking (CDSMD, rec. 57).
[298] See e.g. CDSMD rec. 52: "[...] licensing of rights in audiovisual works to video-on-demand services".
[299] Cf. REGULATION (EU) 2017/1128 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 June 2017
on cross-border portability of online content services in the internal market, rec. 4: "[...] rights for the transmission of content protected by copyright or related rights, such as audiovisual works, are often licensed on a territorial basis"; Dir. 2019/789, rec. 10: "Given the specificities of the financing and licensing mechanisms for certain audiovisual works, which are often based on exclusive territorial licensing [...]".
[300] Reg. 2017/1128, rec.12.
[301] See Art. 3 Directive 2014/26 under (m): 'multi-territorial license' means a license which covers the territory of more than one Member State.
[302] Cf. Rec. 2017/1128, rec. 31: "Contracts under which content is licensed are usually concluded for a relatively long duration."
[303] A. Strowel and B. Vanbrabant, op.cit. 36.
[304] Cf. Art. 16(2) Dir. 2014/26.
[305] Cf. CDSMD rec. 40: "Uses covered by such licenses should not be for profit-making purposes, including where copies are distributed by the cultural heritage institution, such as in the case of promotional material about an exhibition." Cf. also Art. 5(3) Directive 2014/26: "Rightsholders shall have the right to grant licenses for non-commercial uses of any rights, categories of rights or types of works and other subject-matter that they may choose."

(6) Intensity of use (e.g., number of uses/views etc.)

(7) Person of the licensee (e.g., limitation of transfer of the license- in personam license).

c. Further (sub)licensing

Another important variable of a license is the allowance of further granting (sublicensing) of licensed rights by licensees to other persons. Sublicenses are licensing contracts that are signed between the licensee (sub-licensor) and a sublicensee.[306] The limitations of scope in scope and consideration are components equally applicable to sublicenses. However, in case of a sublicense, the sub-licensor acts as licensor vis-à-vis the sublicensee but as a licensee, vis-à-vis the initial licensor. Accordingly, the sublicensee may not grant a sublicense whose scope would exceed the scope of licensed rights. A license which allows for the granting of further sublicenses is usually called "sub-licensable" license.

d. Consideration

License is a reciprocal contract.[307] Usually, a license is granted with consideration. The consideration corresponds to the exchange offered by the licensee for obtaining the authorisation of use or exploitation of licensed content.

Typically, license is granted with **monetary consideration (**license fees), in form of lump sum or royalties payment. Consideration for a license may also be **nonmonetary** (e.g. data or other non monetary consideration).

A license with consideration functions as a vehicle for monetization from a rightsholder (licensor) perspective: the consideration received against the granting of license is a means for monetizing the licensed content.

In the absence of monetary consideration, license is considered as **free/open license**[308] (e.g., Creative commons license).

e. Obligations of the licensee

Another building block for licensing agreements is connected with the usual obligations of the licensee, which mainly refer to:

- specific requirements of crediting (beyond moral rights requirements);
- obligation of exploitation of the licensed content (i.e., obligation of divulgating and disseminating the licensed content);

obligation of transparency/reporting (especially when an obligation of exploitation is stipulated, along with a consideration consisting in royalty payment).

[306] Cf. CDSMD rec. 76: "[...]the rights have been sub-licensed to other parties who exploit the rights [...]".
[307] Opinion V. Trstenjak, 27/1/2009, in the case C-533/07, 23 april 2009, Falco Privatstiftung andThomas Rabitsch vs. Gisela Weller-Lindhorst, para. 51 ('Falco Opinion').
[308] See CDSMD, rec. 82.

17 Annex VIII: Exceptions and Limitations to Copyright

The EU copyright *acquis* harmonises not just a number of economic rights of exploitation, but also provides a number of exceptions and limitations to those rights. In light of the XReco workflow, there are a number that are especially relevant at the stage of input data and content. Limitations and exceptions comprise, in effect, permissions or authorisations for certain actors that are not the author or relevant rightsholder in regard to the work or subject matter at stake, allowing these other actors to perform certain acts that otherwise may infringe on the exclusive rights of the author or rightsholder. It should, however, be noted that some of the exceptions and limitations are not fully harmonised, including because many remain optional for Member States to implement.

A general principle for all exceptions and limitations is the so-called "three-step test". This test emerges from Article 9 of the Berne Convention, stating that: "It shall be a matter for legislation in the countries of the Union to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author." [309] This phrasing has been more-or-less carried over to other international agreements on copyright and related rights.[310] An important caveat is that although the Berne Convention's text refers to "the reproduction of such works", the scope of rights, and thereby the exceptions and limitations to those rights, addressed by the three-step test has been expanded.[311]

The three-step test sets limits to exceptions and limitations on authors' rights.[312] However, the balance between the rights articulated by copyright (and by extension, the EU copyright *acquis*) shift constantly, being "embedded in a complex matrix established by copyright, contract and technical developments" .[313]

In the context of the EU copyright *acquis*, the three-step test is enshrined in the InfoSoc Directive, stating in Article 5 that:[314] "The exceptions and limitations provided for in paragraphs 1, 2, 3 and 4 shall only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightsholder." [315] The applicability of the test has also been extended to new exceptions and limitations introduced after the InfoSoc Directive.[316]

[309]Art. 9(2) BC.

[310]"Members shall confine limitations and exceptions to exclusive rights to certain special cases which do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the rights holder" (Art. 13 TRIPS Agreement); "Contracting Parties may, in their national legislation, provide for limitations of or exceptions to the rights granted to authors of literary and artistic works under this Treaty in certain special cases that do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the author." (Art. 10(1) WIPO Copyright Treaty).

[311]Cf. Bechtold in Dreier and Hugenholtz(eds), 469.

[312]Martin Senftleben, Copyright, Limitations and the Three-Step Test. An Analysis of the Three-Step Test in International and EC Copyright Law(Kluwer Law 2004), 5.

[313]Senftleben 2004, 35.

[314] Prior to the InfoSoc Directive, it was already addressed in regard to the special subject matter of software (Art. 6(3) Software Directive) and databases (Art. 6(3) Database Directive), and for rental and lending rights (Art. 10(3) Rental and Lending Directive); see also Art. 3(3)Directive (EU) 2017/1564 of the European Parliament and of the Council of 13 September 2017 on certain permitted uses of certain works and other subject matter protected by copyright and related rights for the benefit of persons who are blind, visually impaired or otherwise print-disabled and amending Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society [2017] OJ L 242/6 (Marrakesh Directive).

[315] Art. 5(5) InfoSoc Directive.

[316]Namely, regarding the new exceptions and limitations introduced by the CDSM Directive –"Article 5(5) of Directive 2001/29/EC shall apply to the exceptions and limitations provided for under this Title. The first, third and fifth subparagraphs of Article 6(4) of Directive 2001/29/EC shall apply to Articles 3 to 6 of this Directive"(Art. 7 CDSM Directive).

Concretely, the three-step test circumscribes the EU Member States' discretion in formulating and transposing exceptions and limitations as part of the copyright *acquis* while striking a fair balance between EU fundamental rights.[317] The Court of Justice of the European Union has addressed the relevance of the three-step test in concrete cases,[318] it remains an open question whether three-step test as stated in the provision of Article 5(5) InfoSoc Directive is addressed only to the legislature of the Member States, or whether courts of Member States shall apply the test in each of their decisions regarding limitations and exceptions.[319]

17.1 Art. 5(1) InfoSoc Directive

Certain acts envisioned essential to the process of generating XR data and content, including by the XReco platform, may comprise temporary acts of reproduction benefitting from Article 5(1) of the InfoSoc Directive. Under this provision, temporary acts of reproduction which are do not infringe the right of reproduction of the rightsholder at stake where:[320] (1) the act is temporary; (2) it is transient or incidental; (3) it is an integral and essential part of a technological process; (4) the sole purpose of that process is to enable a transmission in a network between third parties by an intermediary or a lawful use of a work or protected subject matter, and; (5) that act does not have any independent economic significance.

The CJEU has further clarified that these criteria are cumulative,[321] and that they must be interpreted strictly.[322] The CJEU has also held that the exception must be interpreted in light of Article 5(5) of the InfoSoc Directive – the so-called three-step test.[323]

Generally, this exception can be understood to imply that acts such as internet browsing and local caching do not interfere with copyright.[324] Specifically, the CJEU has clarified that Article 5(1) InfoSoc Directive must strike a "fair balance between the rights and interests of rights holders and of users of protected works who wish to avail themselves of those technologies" .[325]

17.2 Articles 3 and 4 CDSM Directive: Text and Data Mining

The Copyright in the Digital Single Market Directive (CDSM) introduced two new exceptions in Articles 3 and 4 to the right of reproduction, the right of extraction for *sui generis* databases and for the right of press publishers under Art. 15(1) of the CDSM for "text and data mining" (TDM). As two separate provisions, they have different

[317]Case C-516/17 *SpiegelOnline* [2019] ECLI:EU:C:2019:625, paras. 37-38; Case C-469/17 *Funke Medien* [2019] ECLI:EU:C:2019:623, paras. 52-53.

[318]E.g. finding that on-screen copies and the cached copies made by an end-user in the course of viewing a website fulfils the three-step test in the context of the temporary reproduction exception ("the viewing of websites by means of the technological process at issue represents a normal exploitation of the works which makes it possible for internet users to avail themselves of the communication to the public made by the publisher of the website concerned. Given that the creation of the copies in question forms part of such viewing, it cannot operate to the detriment of such an exploitation of the works." (Case C-360/13*Public Relations Consultants Association* [2014] ECLI:EU:C:2014:1195, para. 61).

[319]Stefan Bechtold, 'Directive 2001/29/EC' in Thomas Dreier and P Bernt Hugenholtz (eds),*Concise European Copyright Law* (2nd ed, Wolters Kluwer 2016), 469.

[320] Art. 5(1) InfoSoc Directive; Case C-302/10 [2012] *Infopaq II* ECLI:EU:C:2012:16, para 54.

[321]*InfopaqI*, para 55; *Infopaq II*, para 26; Case C-527/15 *Filmspeler*[2017] ECLI:EU:C:2017:300, para. 61.

[322]*Infopaq I*, paras. 56-57; Infopaq II, para. 27; Case C-429/08 *Football Association Premier League*[2011] ECLI:EU:C:2011:631, para. 162;*Public Relations Consultants Association*, para. 23; Case C-527/15*Filmspeler*, para. 62.

[323]*Infopaq I*, para. 58; *Filmspeler*, para. 63; see Section **Fehler! Verweisquelle konnte nicht gefunden werden.** above.

[324]Recital 33 InfoSoc Directive.

[325]Public Relations Consultants, para. 24.

scopes and features which need to be considered separately, though they share certain aspects that can be generalised.

Generally, these exceptions provide that, given that a beneficiary has lawful access to the subject matter and requirements of each provision are fulfilled, they can perform TDM. TDM is defined as "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations" (Art. 2(2) CDSM). As Margoni and Kretschmer point out, "[this] certainly includes most modern, data-driven forms of AI, such as traditional machine learning and more advanced forms of deep learning and neural network structures".[326]

There are certain general caveats to the availability of the TDM exceptions:

- The availability of the inputs for TDM may not just depend on the copyright status of the input itself, but also of some underlying object that is contained in the input, such as where a photograph that is a copyright-protected work includes a real-life object that is also a copyright-protected work.
- Not all inputs that may be relevant for the uses of TDM are necessarily covered by copyright or the two applicable related rights. Namely, certain potentially important works or other subject matter, as well as the relevant rights attached to the subject matter in question, may not be addressed by the TDM exceptions. This means for instance that where non-original photographs are protected in a given Member State,[327] these are not addressed by the TDM exceptions. Furthermore, the exceptions do not address the right of communication to the public and the right to make available to the public,[328] nor do they address the right of distribution.[329] The lack of availability of these rights may mean that where the output of the text and data mining technique contains a reproduction, it cannot be distributed or communicated to the public.[330]
- The output of the TDM operation is not necessarily protectable by copyright as a new work. Only where the output is an original work of authorship can this be ascertained.[331] As Hugenholtz and Quintais underline, in order for AI-assisted outputs to qualify for copyright protection under the EU *acquis*, the output must be: "(1) in relation to "production in the literary, scientific or artistic domain"; (2) the product of human intellectual effort; and (3) the result of creative choices that are (4) "expressed" in the output" .[332] In that regard, authorship will be protected where creative contributions have been made,[333] though whether this is achieved by each output depends on a case-by-case analysis.

The rights that the TDM exceptions are exceptions to are not homogeneous. For instance, the right of reproduction for copyright-protected works under Article 2 InfoSoc Directive is strictly speaking different from the right to extract data from a *sui generis* database under Art. 7 Database Directive.[334] Further, for the right of

[326]Margoni and Kretschmer, 688.
[327]As permitted by Art. 6 Term Directive: "Member States may provide for the protection of other photographs."
[328] Art. 3 InfoSoc Directive.
[329]Art. 4 InfoSoc Directive.
[330]Cf. Margoni and Kretschmer, 693.
[331]See Section **Fehler! Verweisquelle konnte nicht gefunden werden.** regarding the requirements for copyright protection.
[332]P Bernt Hugenholtz and João Pedro Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 IIC - International Review of Intellectual Property and Competition Law 1190, 1212.
[333]Hugenholtz and Quintais, 1207.
[334]For instance, whereas the right of extraction covers acts of temporary or permanent transfer (Art. 7(2) Database Directive), it does not cover "consultation of a database" (Case C-203/02 *British Horseracing Board* [2004] ECLI:EU:C:2004:695, para. 54).

reproduction of publishers of press publications,[335] there currently is no established practice for its operationalisation vis-à-vis exceptions and limitations.[336]

The common prerequisite of lawful access is not always straightforward. The notion is addressed in recitals 14 and 18 of the CDSM Directive regarding Articles 3 and 4 respectively, which clarify concrete practices that lawful access is understood to cover, yet recital 14 also mentions "other lawful means" which are not elaborated further, whereas recital 18 clarifies that subject matter is accessed lawfully "including when it has been made available to the public online". While this clarifies certain cases, the boundaries of the notion remain difficult to ascertain.

Finally, it should be noted that given the nature of Directives under EU law, these exceptions are not transposed or implemented in complete uniform fashion across the EU.[337]

Overall, these exceptions may play an important role for the XReco project.[338]

17.2.1 Article 3: TDM for Research Organisations and Cultural Heritage Institutions

Article 3 of the CDSM Directive provides an exception for "reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access" .[339] For the purposes of this provision, research organisations are defined as "a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: (a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation" ,[340] whereas cultural heritage institutions are defined as "a publicly accessible library or museum, an archive or a film or audio heritage institution" .[341] "For the purposes of scientific research" is not explicitly defined, however, regarding "scientific research", it is clarified that it "should be understood to cover both the natural sciences and the human sciences" .[342]

Beyond the delimitation of this exception to research organisations and cultural heritage institutions, this exception has the crucial caveat that the demarcation to private sector entities collaborating with beneficiaries of the exception for the purposes of scientific research can be difficult to ascertain. Specifically, whereas the Directive stipulates that research organisations and cultural heritage institutions "should also be able to rely on their private partners for carrying out text and data mining, including by using their technological tools",[343] it nevertheless highlights that "organisations upon which commercial undertakings have a decisive influence allowing such undertakings to exercise control because of structural situations, such as through their quality of

[335]Which is limited to publications "for their online uses by information society service providers" (Art. 15 CDSM Directive).
[336]Especially as Article 15 CDSM Directive carves out "individual words or very short extracts" from the scope of protection (Cf. Ula Furgał, 'The EU Press Publishers' Right: Where Do Member States Stand?' (2021) 16 Journal of Intellectual Property Law and Practice 887).
[337] Cf. Eleonora Rosati, 'Are Directives Good for the EU Internal Market? The Case of the Copyright DSM Directive and Its National Transpositions' (2021) 16 Journal of Intellectual Property Law and Practice 1027.
[338] Leander Stähler, 'The Production of 3D Digital Assets with NeRF: An Opportunity for the EU TDM Exceptions?' (*CiTiP blog*, 29 September 2023) <https://www.law.kuleuven.be/citip/blog/the-production-of-3d-digital-assets-with-nerf-an-opportunity-for-the-eu-tdm-exceptions/> accessed 30 September 2023.
[339]Art. 3(1) CDSM Directive.
[340]Art. 2(1) CDSM Directive.
[341]Art. 2(3) CDSM Directive.
[342]Recital 12 CDSM Directive.
[343]Recital 11 CDSM Directive.

shareholder or member, which could result in preferential access to the results of the research, should not be considered research organisations for the purposes of this Directive.[344] This arguably imperils the ability of a business acting for non-commercial scientific research purposes to benefit from the exception.[345]

17.2.2 Article 4: General TDM Exception

Article 4 of the CDSM Directive provides and exception for "reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining" .[346] Notably, it not only addresses the horizontal right of reproduction under the InfoSoc Directive, the rights regarding databases and the right of press publishers for online use, but also the right of reproduction and the right of adaptation for the vertical regime for software.[347] This exception is of general applicability, meaning that it is not limited to research organisations or cultural heritage institutions, nor is it limited to the purposes of scientific research.

There are two notable caveats to this exceptions. Firstly, Article 4(2) stipulates that reproductions and extractions "may be retained for as long as is necessary for the purposes of text and data mining". This entails that once automated analytical techniques have been applied to the text and data in question, copies thereof should be discarded.

Secondly, for this exception there exists a so-called "opt-out" under Article 4(3), which states that "shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightsholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online". As some have argued, this makes the availability of the exception unlikely where private ordering and technological protection measures are involved.[348]

17.2.3 Art. 6 CDSM Directive: Digital preservation

Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) of Directive 2009/24/EC and Article 15(1) of this Directive, in order to allow cultural heritage institutions to make copies of any works or other subject matter that are permanently in their collections, in any format or medium, for purposes of preservation of such works or other subject matter and to the extent necessary for such preservation.

17.2.3.1 XR Content Generation

According to the exception introduced by Article 6 of the CDSMD titled 'Preservation of cultural heritage' Member States shall provide for an exception to the reproduction right under copyright and related rights and to the right of extraction under sui generis right in order to allow cultural heritage institutions to make copies of any works or other subject matter that are permanently in their collections, in any format or medium, **for purposes of preservation** of such works or other subject matter and to the extent necessary for such

[344]Recital 12 CDSM Directive.
[345]Margoni and Kretschmer, 695.
[346]Art. 4(1) CDSM Directive.
[347] Article 4(1)(a) and (b) of Directive 2009/24/EC.
[348] J Griffiths et al, 'Comment of the European Copyright Society Addressing SelectedAspects of the Implementation of Articles 3 to 7 of Directive (EU)2019/790 on Copyright in the Digital Single Market'(*European Copyright Society*, 3 May 2022) <https://europeancopyrightsocietydotorg.files.wordpress.com/2022/05/ecs_exceptions_final-3.pdf>, 17-18.

preservation. As 'cultural heritage institution' (CHI) refers mainly[349] to a publicly accessible library or museum, an archive or a film or audio heritage institution (Art. 2 (3) CDSMD).

Given the specific purpose (preservation of works that are permanently in CHI collections, which may, however, be widely interpreted[350]) and the beneficiary (CHI) of the exception, it is not expected to be applied in view of covering any reproductions that take place within the framework of the XReco operation.

17.2.4 Art. 14 CDSM Directive: Works of visual art in the public domain

"Member States shall provide that, when the term of protection of a work of visual art has expired, any material resulting from an act of reproduction of that work is not subject to copyright or related rights, unless the material resulting from that act of reproduction is original in the sense that it is the author's own intellectual creation." (Art. 14 CDSM)

17.2.4.1 XR Content Generation

According to Article 14 CDSM titled "Works of visual art in the public domain", Member States shall provide that, when the term of protection of a work of visual art has expired, any material resulting from an act of reproduction of that work is not subject to copyright or related rights, unless the material resulting from that act of reproduction is original in the sense that it is the author's own intellectual creation. According to this provision, no copyright or related right may be recognized in faithful[351] representations of works of visual arts (such as paintings but also photographs) that are not protected anymore, due to the expiry of term of protection. This may be proven as a useful provision in the XReco context, to the extent that pre-existing content that consists in identical reproduction of public domain works (such digital copies of public domain images or photographs[352]) would not possibly be subject to copyright or any related rights protection under national law of any EU Member state and, therefore, no licensing transactions vis-à-vis the said content would be necessary in any licensing instance (upload, description analysis, reuse etc).

[349] See E. Rosati, Copyright in the Digital Single Market. Article-by-Article Commentary to the Provisions of Directive 2019/790, Oxford University Press, 2021, 134-135.
[350] Rosati, op.cit., p136-137.
[351] I. Stamatoudi andPaul Torremans, The Digital Single Market Directive in I. Stamatoudi/Paul Torremans (ed.), EU Copyright Law. A Commentary, 2nd Edition, (EE 2021), 719.
[352] Rosati, op.cit., 242.

18 Annex IX: Smart Legal Contracts and XR Data and Content

18.1 Smart Contracts and Smart Legal Contracts

Smart legal contracts are tools that aim to combine the utility of smart contracts and contracts in a legal sense.[353] In forming an appraisal of smart legal contracts, within the context of EU law, there are two dimensions to be addressed. First, smart contracts as such, that is, certain computer programs or similar digital functionalities as technologies. Second, the validity of such smart contracts in the context of legal relations, namely as a contract or a license.

In order to address smart contracts as a category of technologies, it is necessary to define how they are understood, and specifically how they are described in the context of XReco.

Smart contracts as an application of blockchain or distributed-ledger technology (DLT) emerged with the usage of the Ethereum blockchain. The developer of the Ethereum blockchain, Vitalik Buterin, described smart contracts as an important inquiry in 2014, defining "smart contracts" as "systems which automatically move digital assets according to arbitrary pre-specified rules" .[354] In this definition "[anyone] can create their own arbitrary rules for ownership, transaction formats and state transition functions,"[355] effectively, "cryptographic "boxes" that contain value and only unlock it if certain conditions are met" .[356]

In a general sense, blockchain technology constitutes an infrastructure for the storage of data and the management of software applications,[357] a database that is replicated across a network of computers updated through a "consensus algorithm" .[358] De Filippi and Wright argue that blockchain-based technologies "have the capacity to implement their own system of rules", that are enforced the underlying protocol and smart contracts.[359] They find that "Traditional legal doctrines, especially those focused on regulating middlemen, will not easily translate to these new decentralized and autonomous systems" .[360]

De Filippi and Wright's discussion builds upon the arguments put forward by Reidenberg and Lessig regarding *lex informatica* and (software) codes as law respectively,[361] highlighting that code is one of multiple regulatory factors exerting a normative influence on behaviour.[362] In light of this, code can take a dominant position in

[353]Which can be defined through their characteristics of automatability and enforceability (Christopher D Clack, Vikram A Bakshi and Lee Braine, 'Smart Contract Templates: Foundations, Design Landscape and Research Directions' (arXiv, 15 March 2017) <http://arxiv.org/abs/1608.00771>).

[354]Vitalik Buterin, 'Ethereum: A Next-Generatiion Smart Contract and Decentralised Application Platform' (*Ethereum*, 2014) <https://ethereum.org/669c9e2e2027310b6b3cdce6e1c52962/Ethereum_Whitepaper_-_Buterin_2014.pdf>.

[355]Buterin, 13.

[356]Buterin, 13.

[357]Primavera De Filippi and Aaron Wright, *Blockchain and the Law: The Rule of Code* (First Harvard University Press paperback edition, Harvard University Press 2019), 33

[358] Michèle Finck, *Blockchain Regulation and Governance in Europe* (1st edn, Cambridge University Press 2018., 1; refined as "a shared and synchronized digital database that ismaintained by an algorithm and stored on multiple nodes (the computers that storea local version of the distributed ledger)" (Finck, 6).

[359]De Filippi and Wright dub this phenomenon "lex cryptographica" (De Filippi and Wright, 50)

[360]De Filippi and Wright, 52.

[361]Joel R Reidenberg, 'Lex Informatica: The Formulation of Information Policy Rules through Technology' (1997) 76 Texas Law Review 553;Lawrence Lessig, *Code* (Version 2.0, Basic Books 2006).

[362]Finck, 39.

regulating particular behaviour, even leading to the avoidance of legal regulation, but it can also be constrained or supported by the law[363] – in essence a dialogic relationship.[364]

At a formal legal level, several critical responses to the usage of smart contracts have been put forward. For one, Eliza Mik highlights, smart contracts that are self-enforcing and tamper-proof can create a "cascade of problems" :[365]

- If a smart contract is understood to also comprise **contractual performance**, it must be ensured that the code of this smart contract contains no errors, or failing that, to allocate the risk of the occurrence of malfunctions by prior agreement.[366]
- If the smart contract is **self-enforcing**, it is necessary to ensure that it self-enforces as intended or promised, reflecting the parties' agreement.[367]
- If the smart contract is **tamper-proof**, it is necessary that all possible events that may occur during its lifetime that affect its operation be anticipated.[368]

Regarding the relationship of smart contracts to contractual performance, it is important to note that applications of smart contracts do not necessarily comprise contractual performance as such,[369] but rather it can be explored whether aspects of contractual performance can occur through the format of a smart contract. Further, regarding the self-enforcing nature of smart contracts, they can be designed such that certain aspects of the smart contract upon which the parties' univocally agree upon (e.g., on access to a particular piece of content or data) may be self-enforcing, but not necessarily all aspects. Regarding the tamper-proof nature of smart contracts, and relatedly, they can be designed such that some aspects are tamper-resistant, while also providing tools to change those aspects. Further, where not all aspects of the smart legal contract are technologically-mediated through a smart contract, there may simply not be a need for the entirety of the smart legal contract to be expressed at the technological level. Furthermore, Karen Levy's scrutiny of smart contracts as an "automated enforcement framework" highlights certain contracting practices in which people engage to manage their relations that pursue important social aims.[370] This automated enforcement framework is understood to pursue the aim to "collapse contract formation and enforcement into a single instrument" ,[371] highlighting that such blockchain-based solutions focus on *ex ante* measures at the expense of *ex post* ones.[372]

18.2 Smart Legal Contract

Some of the aspects of smart contracts, particularly as they have emerged as a concept in the discussion of blockchain-based smart contracts, can be reconciled. Clack et al demarcate *smart legal contracts*– where the

[363]Finck, 39-43
[364] Margoni and Kretschmer; see also Finck, 44.
[365]Eliza Mik, 'Smart Contracts: Terminology, Technical Limitations and Real World Complexity' (2017) 9 Law, Innovation and Technology 269, 281.
[366]Mik, 281.
[367]Mik, 282.
[368]Mik, 282 et seq.
[369]Mik's analysis is based on an stringent understanding of smart contracts, meaning that the risk management strategies proposed are equally stringent. This understanding can be expanded in light of the smart legal contract notion, such that these problems and the risks they pose are addressed proactively.
[370]Karen EC Levy, 'Book-Smart, Not Street-Smart: Blockchain-Based Smart Contracts and The Social Workings of Law' (2017) 3 Engaging Science, Technology, and Society 1, 5.
[371]Levy, 3.
[372]Levy 2017, 4.

agreement is a legal agreement, at least some of which is capable of being implemented in software – from *smart contract code* – which is automated software that may not necessarily be linked to a formal legal agreement.[373]

This taxonomy creates a number of distinctions usable for XReco. Clack et al highlight that some aspects may be automatable, but not necessarily automated.[374] In line with this, they highlight that some aspects may be enforceable, but necessarily enforced at the network level via "tamper-proof technology" .[375] This ultimately means that one can scrutinise between what they call "operational aspects" of any given contract, that is, the parts of the contract that "we wish to automate", and "non-operational aspects", the parts of the contract that "we do not with to (or cannot) automate" .[376]

This implies the following:[377]

Type	Smart contract code	Smart legal contract
Characteristics	- Concerned with operational aspects expressed in code - Focused exclusively on automated parts	- Concerned with both the operational and non-operational aspects of a contract - Aspects of the contract may be automated

18.3 Smart contracts in EU law

In general, electronic documents have legal effect.[378] Electronic documents are defined as "any content stored in electronic form, in particular text or sound, visual or audiovisual recording" .[379] "An electronic document shall not be denied legal effect and admissibility as evidence in legal proceedings solely on the grounds that it is in electronic form."[380]

There may be further applicable rules for: (1) Smart contracts that are used to execute data sharing agreements; (2) Smart contracts that protect copyright and related rights-protected content; (2) Smart contracts that are based on an "electronic ledger".

The Data Act includes "[essential] requirements regarding smart contracts for executing data sharing agreements".[381] Data is defined in the Data Act as "any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audio-visual recording",[382] whereas a smart contract is defines as "a computer program used for the automated execution of an agreement or part thereof, using a sequence of electronic data records and ensuring their integrity and the accuracy of their chronological ordering".[383] Where these definitions apply to the smart contract in questions, "[the] vendor of an application using smart contracts or, in the absence thereof, the person whose trade, business or profession involves the deployment of smart contracts for others in the context of executing an

[373]Clack et al, 2.
[374]Clack et al, 3.
[375]Clack et al, 4.
[376]Clack et al, 5.
[377]Adapted from Clack et al 2017, 6.
[378]Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC[2014] OJ L 257/73 (eIDAS Regulation).
[379]Art. 3(35)eIDAS Regulation.
[380]Art. 46eIDAS Regulation.
[381]Art. 36 Data Act.
[382]Art. 2(1) Data Act; cf. Art. 2(1) DGA.
[383]Art. 2(39) Data Act.

agreement or part of it, to make data available shall ensure that the smart contracts comply with [the essential requirements]".[384] These essential requirements are:

- Robustness and access control;[385]
- Safe termination and interruption;[386]
- Data archiving and continuity;[387]
- Access control,[388] and;
- Consistency.[389]

More specifically, the InfoSoc Directive provides for MSs to provide "adequate legal protection against the circumvention of any effective technological measures (…)".[390]In this provision "technological measures" is defined as "any technology, device or component that, in the normal course of its operation, is designed to prevent or restrict acts, in respect of works or other subject-matter, which are not authorised by the rightsholder of any copyright or any right related to copyright as provided for by law or the sui generis right provided for in Chapter III of Directive 96/9/EC".[391] Whereas it is deemed to be "effective", where "the use of a protected work or other subject-matter is controlled by the rightsholders through application of an access control or protection process … which achieves the protection objective".[392] Where smart contracts fulfil this definition, they may be protected against circumvention.[393]

It should be noted that the notion of smart contract under the Data Act is technology-neutral,[394] but it can be connected what is referred to as an "electronic ledger".[395] Electronic ledgers are currently not explicitly regulation, but the 2022 Commission eIDAS Regulation Amendment Proposal would add "electronic ledgers" to the catalogue of "trust services". Under this proposal, an electronic ledger is defined as "a tamper proof electronic record of data, providing authenticity and integrity of the data it contains, accuracy of their date and time, and of their chronological ordering".[396] According to the proposal, these electronic ledgers shall not be denied legal effect "solely on the grounds that it is in an electronic form or that it does not meet the requirements

[384]Art. 36(1) Data Act.

[385] "ensure that the smart contract has been designed to offer access control mechanisms and a very high degree of robustness to avoid functional errors and to withstand manipulation by third parties" (Art. 36(1)(a) Data Act).

[386]"ensure that a mechanism exists to terminate the continued execution of transactions and that the smart contract includes internal functions which can reset or instruct the contract to stop or interrupt the operation, in particular to avoid future accidental executions" (Art. 36(1)(b) Data Act).

[387]"ensure, in circumstances in which a smart contract must be terminated or deactivated, there is a possibility to archive the transactional data, smart contract logic and code in order to keep the record of operations performed on the data in the past (auditability)" (Art. 36(1)(c) Data Act).

[388] "ensure that a smart contract is protected through rigorous access control mechanisms at the governance and smart contract layers" (Art. 36(1)(d) Data Act).

[389]"ensure consistency with the terms of the data sharing agreement that the smart contract executes" (Art. 36(1)(e) Data Act).

[390]Art. 6(1) InfoSoc Directive, Cf. Software Directive and Conditional Access Directive.

[391]Art. 6(3) InfoSoc Directive.

[392]Art. 6(3) InfoSoc Directive.

[393]Michèle Finck and Valentina Moscon, 'Copyright Law on Blockchains: Between New Forms of Rights Administration and Digital Rights Management 2.0' (2019) 50 IIC - International Review of Intellectual Property and Competition Law 77.

[394]Recital 104 Data Act.

[395] "The notion of "smart contract" in this Regulation is technologically neutral. Smart contracts can, for instance, be connected to an electronic ledger." (Recital 80 Data Act).

[396]Art. 3(53) Commission, 'Proposal for a Regulation of the European Parliament and of the Council amending Regulation (EU) No 910/2014 as regards establishing a framework for a European Digital Identity' COM(2021) 281 final (eIDAS Amendment Proposal).

for qualified electronic ledgers".[397] The proposal also stipulates requirements for qualified electronic ledgers stipulated.[398] Such qualified electronic ledgers would fall under the general scheme for qualified trust services.[399]

Regime	Regulatory Object	Definition	Requirements
Copyright and Related Rights	Technological Protection Measures	Art. 6(3) InfoSoc Directive	Art. 6(4) regarding certain exceptions and limitations + *Nintendo* case
eIDAS [+proposed changes]	(1) Electronic documents [proposed] Electronic ledgers	Art. 3 eIDAS Regulation (1) para. 35 [proposed] para. 53	[Proposed] for qualified electronic ledgers
Data Act (not yet in force)	Smart contracts for executing data sharing agreements	Art. 2 (39)	Art. 36

[397]Art. 45heIDAS Amendment Proposal.
[398]Art. 45ieIDAS Amendment Proposal.
[399]Chapter III eIDAS Regulation.