Enhancing and Re-Purposing TV Content for Trans-Vector Engagement (ReTV)

H2020 Research and Innovation Action - Grant Agreement No. 780656

Enhancing and Re-Purposing TV Content for Trans-Vector Engagement

Deliverable 3.1 (M10)

Metadata and Viewer Profiling

Version 1.0

This document was produced in the context of the ReTV project supported by the European Commission under the H2020-ICT-2016-2017 Information & Communication Technologies Call Grant Agreement No 780656

DOCUMENT INFORMATION

Delivery Type	Report
Deliverable Number	3.1
Deliverable Title	Metadata and Viewer Profiling
Due Date	M10
Submission Date	October 31, 2018
Work Package	WP3
Partners	MODUL Technology, Genistat
Author(s)	Lyndon Nixon (MODUL), Krzysztof Ciesielski (Genistat)
Reviewer(s)	Basil Philipp (Genistat)
Keywords	Metadata interoperability, Viewer profiling
Dissemination Level	PU
Project Coordinator	Vrije Universiteit Amsterdam De Boelelaan 1081 , 1081 HV, Amsterdam, The Netherlands
Contact Details	Coordinator: Prof Lora Aroyo (lora.aroyo@vu.nl) R&D Manager: Dr Lyndon Nixon (lyndon.nixon@modultech.eu) Innovation Manager: Bea Knecht (bea@zattoo.com)

Revisions

Version	Date	Author	Changes
0.1	11/9/18	Krzysztof Ciesielski	Created template and ToC
0.2	27/9/18	Lyndon Nixon	First draft of section on metadata interoperability
0.4	5/10/18	Krzysztof Ciesielski	First draft of section on user profiling
0.6	10/10/18	Lyndon Nixon	Completed section on metadata interoperability
0.8	15/10/18	Krzysztof Ciesielski Lyndon Nixon	Completed first draft for QA
0.9	19/10/18	Basil Philipp	QA review
1.0	23/10/18	Krzysztof Ciesielski Lyndon Nixon	Updates according to the QA
1.1	29/10/18	Lyndon Nixon	Final Check
1.2	30/10/18	Basil Philipp	Last updates according to the final check

Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

This deliverable reflects only the authors’ views and the European Union is not liable for any use that might be made of information contained therein.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In the ReTV project, the objective of our work in WP3 is to deliver components for content adaptation, re-purposing, scheduling and recommendation based on the annotations, metrics and predictions of WPs 1 and 2. To support viewer-level recommendation and audience targeting, we define viewer profiles (T3.2). Since the data produced in the different WPs may follow different metadata models and vocabularies, we also ensure interoperability through the definitions of mappings of properties and values across metadata specifications (T3.1). This deliverable reports on the outputs of these first two tasks, which are a precondition for the subsequent work in implementing re-purposing, recommendation and scheduling components. The remaining tasks of WP3, T3.3 and T3.4, form integral parts of the TVP and will be used in all of the scenarios defined in WP5 and WP6 where they will enable the automatic creation of summaries and find the optimal publishing strategies for content.

1 Introduction

This document summarizes the current state of the WP3 tasks in the first phase of ReTV project (month 10). Chapter 2 discusses task T3.1 (*Metadata and Vocabulary Interoperability*) led by MODUL Technology. Chapter 3 discusses task T3.2 (*Viewer Profiling*) led by Genistat. The outputs of both tasks are preconditions for the progress in the tasks on content re-purposing, recommendation and scheduling. Chapter 4 contains conclusions and the future outlook for all WP3 tasks.

2 Metadata and Vocabulary Interoperability

Metadata and vocabulary interoperability is important within any industry and for any software tool that wishes to be used widely by different organisations. We address the issue of data interoperability in ReTV through (a) re-use of industry standards and specifications where appropriate, and, in the case of the development of our own specifications to meet our own requirements, (b) definition of mappings between our specification and other industry standards and specifications.

2.1 Problem Statement

Heterogeneity in both the metadata models used in multimedia and Web annotation as well as the vocabularies used to refer to values in those metadata models can be a barrier to uptake for the Trans-Vector Platform. After all, any media organisation seeking to make use of the ReTV services in order to acquire richer metrics and data analyses for their content, predictive analytics, or suggested content for (re)publication will want to integrate those services with their existing media asset management systems (MAMS), content management systems (CMS) and/or other existing software and platforms, where their content is already referenced (ID) and described (annotation) in some manner. A lack of interoperability between data models and vocabularies used by the organization and those used in ReTV for the Trans-Vector Platform and its components/services can lead to additional costs for the organization, complicate or discourage selection of ReTV tools/services and fundamentally limit future uptake and exploitation of the project results.

2.2 State-of-the-Art Survey

Heterogeneity in metadata models and vocabularies is par for the course when it comes to the media industry - many are using their own internal specifications, others make use of standardized models in limited or generic ways (e.g. Dublin Core properties). There are efforts from various sides to provide useable standards, both for the media/TV industry specifically and for online (web) content generally, thinking primarily of the EBU (European Broadcasters Union) for the former and the W3C(World Wide Web consortium) for the latter. Considering the different areas of content description used in ReTV, we present here briefly the primary standards/specifications for each area (whether a metadata model or a vocabulary):

- Metadata models for content description
- Vocabularies for semantic annotation
- Vocabularies for visual concepts
- Vocabularies for content categorization

2.2.1 Metadata models for content description

W3C Ontology for Media Resources[1]

The Ontology for Media Resources is a core vocabulary of descriptive properties for media resources, created by the W3C. It is a W3C recommendation since February 2012, produced by the Media Annotations Working Group and provides mapping tables for metadata from many other standards. Its main goal is to bridge the different descriptions of media resources and provide a coherent set of media metadata properties along with their mappings to existing metadata standards and formats. The Ontology for Media Resources provides also implementation compatible with Semantic Web paradigm in RDF/OWL form. As a W3C specification, there is a focus on re-use of the Web architecture and other specifications in this model, e.g. URLs/URIs as property values.

W3C Web Annotation Model[2]

The Web Annotation Model aims at developing an open common specification for annotating digital resources, more generic than 'media resources' (images, videos). Therefore it can capture the generic properties of an annotation of any resource, and be extended with media-specific annotations according to the above Ontology for Media Resources when the target of the annotation is digital media - an image or video, for example. The core model consists in fact of one class and two relations:

– oa:Annotation: The class for Annotations.

– oa:hasBody: The relationship between an Annotation and the Body of the Annotation

– oa:hasTarget: The relationship between an Annotation and the Target of the Annotation

Schema.org[3]

The lack of a common 'standard' for describing the content of Web pages in a structured form, linked to semantic knowledge about the content, which could be used by search engines to provide richer, more meaningful results led to the development of a Web 'schema' (a metadata model made up of classes of things and their properties). It is a collaborative, community activity. The target of the annotation is generally assumed to be a Web page, and the body may be related to any type of content that can appear on a Web page. The focus is on content that needs to be better searched, e.g. Events in schema.org assume popular events such as concerts rather than a description of historical events. However, TV related content is part of the schema.org specification, see for example the definition of a TV Series: https://schema.org/TVSeries.

EBU Core[4]

The EBU (European Broadcasting Union) is the collective organization of Europe's 75 national broadcasters claiming to be the largest association of national broadcasters in the world. EBU's technology arm is called EBU Technical. EBU represents an influential network in the media world. The EBU projects on metadata are part of the Media Information Management (MIM) Strategic Programme. MIM benefits from the expertise of the EBU Expert Community on Metadata (EC-M), participation to which is open to all metadata experts, or users and implementers keen to learn and contribute.

The EBUCore (EBU Tech 3293) is the main result of this effort to date and the flagship of EBU's metadata specifications. It can be combined with the Class Conceptual Data Model of simple business objects to provide the appropriate framework for descriptive and technical metadata for use in Service Oriented

[1] https://www.w3.org/TR/mediaont-10/
[2] https://www.w3.org/TR/annotation-model/
[3] https://schema.org
[4] https://tech.ebu.ch/MetadataEbuCore

Architectures. It can also be used in audiovisual ontologies for semantic web and Linked Data environment. EBUCore has high adoption rate around the world. It is also referenced by the UK DPP (Digital Production Partnership). All EBU metadata specifications are coherent with the EBU Class Conceptual Data Model (CCDM). EBUCore is the foundation of technical metadata in FIMS 1.0 (Framework for Interoperable Media Service). FIMS is currently under development. It embodies the idea of sites like Google, Twitter, YouTube and many other websites offer service interfaces to remotely initiate an action, export data, import a file, query for something, etc. FIMS specifies how media services should operate and cooperate in a professional, multi-vendor, IT environment – not just through a website interface. EBUCore is also the metadata schema of reference in the project EUScreen which delivers linked data to Europeana using EBUCore’s RDF/OWL representation. EBUCore has been also published as AES60 by the Audio Engineering Society (AES)14. The W3C Media Annotation Ontology is based on EBU’s Class Conceptual Data Model and is fully compatible with EBUCore which mapping has been defined and published as part of the W3C specification (Figure 1).

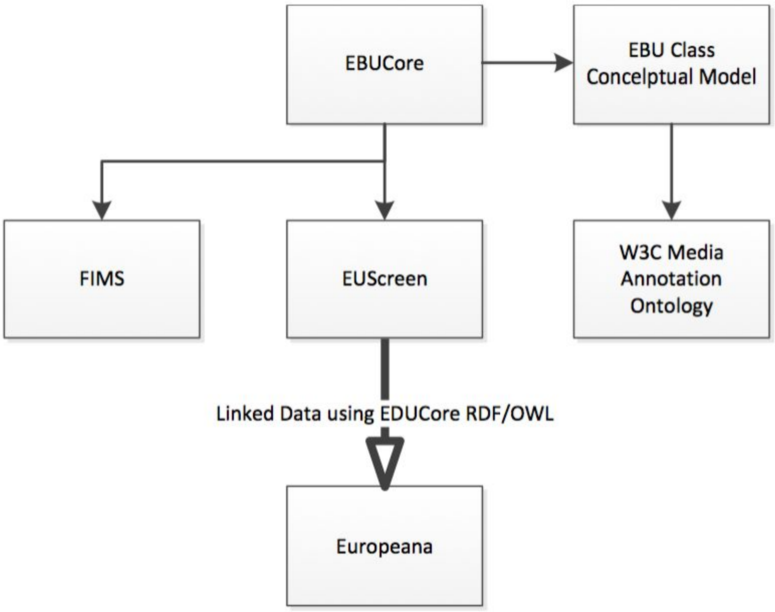

EBUCore
EBU Class Concelptual Model
FIMS
EUScreen
W3C Media Annotation Ontology
Linked Data using EDUCore RDF/OWL
Europeana

Figure 1. Relationship of EBUCore to other metadata models (courtesy LinkedTV Deliverable 2.2)

tvAnytime[5]

The TV-Anytime Forum is a global association of organizations founded in 1999 in USA focusing on developing specifications for audio-visual high volume digital storage in consumer platforms (local AV data storage). These specifications for interoperable and integrated systems should serve content creators/providers, service providers, manufacturers and consumers. The forum created a working group for developing a metadata specification, so-called TV-Anytime and composed of:

– Attractors/descriptors used e.g. in Electronic Program Guides (EPG), or in web pages to describe content (information that the consumer – human or intelligent agent – can use to navigate and select content available from a variety of internal and external sources).

– User preferences, representing user consumption habits, and defining other information (e.g. demographics models) for targeting a specific audience.

[5] https://www.etsi.org/technologies-clusters/technologies/broadcast/tv-anytime

– Describing segmented content. Segmentation Metadata is used to edit content for partial recording and non-linear viewing. In this case, metadata is used to navigate within a piece of segmented content.

– Metadata fragmentation, indexing, encoding and encapsulation (transport-agnostic).

TV Anytime employs the MPEG-7 Description Definition Language (DDL) based on XML to be able to describe metadata structure and also the XML encoding of metadata. TV-Anytime also uses several MPEG-7 datatypes and MPEG-7 Classification Schemes.

BMF[6]

The Broadcast Metadata Exchange Format Version 2.0 (BMF 2.0) has been developed by IRT in close cooperation with German public broadcasters with a focus on the harmonization of metadata and the standardized exchange thereof. The standard particularly reflects the requirements of public broadcasters. BMF contains metadata vocabulary for TV, radio and online content and defines a standardized format for computer-based metadata exchange. It facilitates the reuse of metadata implementations and increases the interoperability between both computer-based systems and different use case scenarios. BMF enables to describe TV, radio and online content as well as production, planning, distribution and archiving of the content. Metadata in BMF are represented in XML documents while the structure for the XML metadata is formalized in an XML Schema. The latest version of the format is BMF 2.0.

egtaMETA[7]

egtaMETA was developed by the EBU specifically for the exchange of information about advertising material. It was published in 2010 (v1.0) in association with EGTA - the association of television and radio sales houses. It is based on EBU Core and defined by an XML Schema. It is primarily a set of semantically defined attributes considered to sufficiently describe advertising material and clustered as follows:

◦ descriptive information (e.g. the title of the advertising spot);

◦ exploitation information (e.g. what is the period during which it shall be used);

◦ credits (inc. keys persons and companies involved in the creation, post-production and release of the advertising spot);

◦ technical information (about the file format and its audio, video and data components).

Europeana Data Model (EDM)[8]

EDM is an XML Schema designed to provide a standard metadata description of cultural heritage objects across a wide range of cultural heritage institutions. As such, it is a specification used for metadata interchange also between archives and audiovisual collections, e.g. NISV uses it in the annotation of their archived TV and film content. It also provides mapping guidelines for these organisations, since each tends towards usage of their own metadata models and vocabularies internally.

2.2.2 Vocabularies for semantic annotation

When providing values for properties in metadata, the original approach has been to use natural language (strings), which led to issues about both human and machine understanding in the future use of that metadata. This led to the development of controlled vocabularies to restrict the values of given properties to a predefined list, where each list item may have an additional definition in the vocabulary that a human user could refer to. With the emergence of knowledge representation, especially on the Web (the "Semantic Web"), these vocabularies evolved into more structured forms from taxonomies through to

[6] http://bmf.irt.de/en

[7] https://tech.ebu.ch/docs/tech/tech3340.pdf

[8] https://pro.europeana.eu/resources/standardization-tools/edm-documentation

ontologies based on different types of formal logics. Even then, they were largely stored away within the organisations that used them, limiting external understanding of the annotations. Finally, public vocabularies emerged on the Web, and based around the idea of Linked Data[9] used also URIs as identifiers for their terms, which meant both humans and machines could look up the additional information about any term using standard Web technologies (HTTP, HTML, RDF). Such vocabularies offer broadly understood terms to be used in any annotation.

DBPedia[10]

The centre of the "Linked Open Data" graph, which illustrates the different publicly available linked data vocabularies, DBPedia is semi-automatically generated out of the structured data (infoboxes and articles) in Wikipedia. As such, almost every Wikipedia article may be represented as a linked entity in DBPedia with an URI as a unique identifier. Entities are also classified, i.e. linked to classes of things which are formalized in their own classes ontology. It is the largest Linked Open Data dataset with more than 4 million entities - however, it is not updated so often (every 6-12 month) and is only as accurate as the data given in Wikipedia. While still the first choice in most semantic annotation, due to the use of DBPedia identifiers in the majority of semantic NER/NEL tools (including our own RECOGNYZE) and the coverage of the dataset, it is also the subject of discussion in much research which relies on DBPedia annotations, due to the perceived messiness of the metadata that is produced (e.g. entity classification is seen as often inconsistent).

WikiData[11]

A community effort to create a Linked Open Data dataset which is cleaner than the automatically produced DBPedia metadata, WikiData is a collaboratively edited knowledge base hosted by the WikiMedia foundation. It is made up of items, which represent entities, and can be described by key-value pairs. It is also sourcing its initial data from Wikipedia, however, a community of users maintain the metadata just like how information is collaboratively created in Wikipedia.

GTAA - Thesaurus for Audiovisual Archives[12]

GTAA is an acronym in Dutch for a 'common thesaurus for audiovisual archives', which is used by NISV to index their archived audiovisual material. The thesaurus consists of several facets for describing TV programs: subjects; people mentioned; named entities (Corporation names, music bands etc); locations; genres; makers and presentators. The GTAA contains approximately 160.000 terms: ~3800 Subjects, ~97.000 Persons, ~27.000 Names, ~14.000 Locations, 113 Genres and ~18.000 Makers, and is continually updated as new concepts emerge on TV. It does follow Linked Data principles, i.e. terms are identified using URIs, and requesting those URIs can retrieve both human or machine readable representations of descriptions of that term.

2.2.3 Vocabularies for visual concepts

Concept vocabularies are based on the work of our partner CERTH through its involvement in the annual TRECVID workshops, as this work has established the training of its visual concept classifiers (Markatopoulou, 2016).

TRECVID-345

TRECVID is an annual series of workshops for the evaluation of different information retrieval tasks in the context of content-based retrieval of digital video. Recent workshops have included tasks for concept

[9] http://linkeddata.org/

[10] http://dbpedia.org

[11] https://wikidata.org

[12] http://gtaa.beeldengeluid.nl/

detection in video (fragments) and have specified the vocabulary of concepts to be supported by systems participating in the task. Our partner CERTH uses the TRECVID-2010 visual concepts list, which contains 500 different concepts but classifiers have been trained on a subset of 345 concepts (the others were discarded due to a lack of sufficient training images).

Places-205

Another vocabulary provided by TRECVID in its 2016 edition was focused on the visual recognition of different types of places, where CERTH trained their classifiers of a selection of 205 concepts from the larger Places dataset (Zhou, 2014).

Other

In recent years, various Internet companies have made available Web services for image and video description, where provided digital media may be labelled automatically with the concepts which appear within them through the use of APIs. While these could provide alternatives for visual concept vocabularies, a study of their media annotations for an e-tourism case noted that they are not standardized nor published online (Nixon, 2018), and thus, unfortunately, each service just contributes to a new heterogeneity in media annotation (and they provide string labels, not Linked Data URIs). The same study provides its own focused annotation vocabulary based on the properties of Destination Image in tourism research, and mappings to that vocabulary from TRECVID and Places vocabularies as well as the IBM Watson visual recognition service. However, this work is too specific (e-tourism) for general re-use.

2.2.4 Vocabularies for content categorization

Content recommendation has generally been based on matching user interests with content which is relevant to those interests. The definition of interests itself can have any level of granularity. In the TV domain, Genre is perhaps the highest granularity and is in common use, e.g. we do get this information in the EPG data. However, a more effective recommendation also needs lower granularity, i.e. a larger range of interest categories, possibly organised in a tree or graph representation so that relevance between different interests can also be calculated (e.g. using various distance algorithms).

tvAnytime / EBU Genre CS[13]

Defined as MPEG-7 Classification Schemes (represented in XML), EBU's Genre CS is a superset of the tvAnytime CS. It contains 1313 Terms in a tree structure with a depth up to 6, although the depth of 4 seems specific enough, e.g. term 3.6.4.6 refers to Folk music or term 3.2.5.6 refers to the sport of Sumo wrestling. The classification scheme is also taken up within EBU Core.

IPTC Subject Codes / Media Topics[14]

Media Topics supercedes the original IPTC subject codes work, which was focused on the description of the subjects of news articles. Media Topics has generalised the usage of the vocabulary, mentioning also the case of digital asset management. It is a 1100 Term taxonomy at a depth of 5.

2.3 RETV APPROACH

We define a shared metadata model and vocabulary for the annotation of content (TV or otherwise) across Web and TV vectors. A conceptual model for ReTV content annotation has been presented in Deliverable 1.1, with a mapping to the document model used by the software platform which will form the basis of our

[13] https://www.ebu.ch/metadata/cs/ebu_ContentGenreCS.xml
[14] http://cv.iptc.org/newscodes/mediatopic

Trans-Vector Platform. Here we provide additional mappings to the "state of the art" annotation models presented above.

Italics are used to indicate that the target property is actually more general than the usage of the property in the ReTV annotation model. Bold is used to indicate the target property is actually more specific than the usage of the property in the ReTV annotation model. Some models re-use properties from the generic Dublin Core Metadata Set, indicated by the prefix dc: or dcterms:.

ReTV Ann. Model	**W3C Media**	**W3C Web (Target)**	**Schema.org (TVSeries)**	**EBU Core**	**tv-Anytime (Content Description)**	**BMF**	**egta-META (spot)**	**EDM**
Genre	genre		genre	genre/ @type Link	Genre	ContentGenre		
Brand			identifier		EpisodeOf	Brand Main Title		
Season			Contains Season					
Episode			episode		**programId** (CRID)			
See also	*relation*		sameAs	*Relation Link*	Related Material	Related Content Instance		dc: relation
Language	language	language	In Language	dc: language	Language	Used Lang-uage	lang-uage	lang-uage
Title	title		name	dc:title	Title	MainTitle	title	dc:title
Descript-ion	descript-ion		descript-ion	dc: descript-ion	**Synopsis**	Program Descript-ion	descri-ption	dc: descript-ion
Keyword	keyword		keywords	*dc:subject*	Keyword	Controlled Keyword, UncontrolledKeyword	subject	dc:subject
Duration	duration		time Required	duration	Published Duration	Broadcast Duration	format / dura-tion	dcterms: extent
Fragment	fragment	Id + URL fragment		hasPart	use the Segment Information Type	Broadcast Segment		dcterms: hasPart
Sentence					use Related	*Script*		

					Material			
Publication/ Locator	locator		url	Format/ Location	**Program URL**	*FileLocator*		
Publication/ Service	*publisher*		provider	*dc: publisher*	serviceID Ref	Publicat-ionService		dc: publisher
Publication/ Channel	*creator*	*creator*	publisher	*creator*				dc:creator
Publication/ start, end			startDate, endDate		Published StartTime, Published EndTime	Broadcast StartDate AndTime		
Is Version of	*relation*		**trailer** example OfWork	*Relation Link*	Derived From	Related Version		

Regarding the W3C Media Ontology, it can be seen that it is designed to annotate (online) digital media and not specifically TV programs, hence there are no equivalent for Brand, Episode and Series (as discussed in deliverable D1.1, we adopted these properties from the BBC Programmes Ontology) nor for publication time (only creation time). Also, there is no concept of textual transcripts to accompany the video data. In fact, this lack of TV focus can also be seen in that only EBU Core is mapped to the model in the W3C specification, mappings to the other models - from the TV or archive industries - are not defined. Schema.org's TVSeries class covers almost all of our annotation model, providing for some interpretation of the properties (which come from the more general CreativeWork) but only misses references to fragments (of video or to video transcripts). tvAnytime covers the properties of a TV programme including its broadcast at a certain time, but its use of XML and MPEG-7 makes the modelling of the metadata more complex. It also ties program identifier to its own Content Reference IDs (CRIDs), and so is tied to TV content and not re-usable for Web content which uses URLs. BMF, as an industry metadata format, distinguishes the description of media assets at different stages in the lifecycle (Production-Editorial-Publishing) which also makes direct mappings more difficult - the Package for "Publishing", for example, defines the BroadcastScheduleSet (which includes the BroadcastStartDateAndTime and BroadcastDuration as well as the PublicationService) but the program description, brand title etc. each occur in different other Packages, reflecting BMFs purpose for metadata exchange about the internal state of a media asset at different stages of its lifecycle, whereas reTV models a full description of the program at the time of publication. egtaMETA is specific to the description of advertisements, whereas we consider a TVAd as a specialisation of the TVProgram class. As can be seen, ads share many properties of programs (title, description...). Mainly, egtaMETA lacks any means to capture relationships between ads (as with our "isVersionOf") or individual publications of an ad (the PublicationHistory mainly captures the first transmission) which is necessary for associating different metrics with each publication. It adds properties to describe the product and brand being advertised, while we have a single property to add the label of the brand detected visually in the video (cf. deliverable D1.1 Brand Detection). EDM is likewise focused on the annotation of individual media assets which are being archived rather than published (repeatedly), and thus lacks metadata properties for Brand/Season/Episode or for Publication events.

We have also proposed which vocabularies to use for certain metadata properties, with a focus on re-use of Linked Open Data URIs for the semantic annotation, concept labels from controlled vocabularies known to the research community for visual concept detection, and a fixed list of content categories for recommendation.

For semantic annotation, providing an URI as a value for a property is a standard approach with the approach of Linked Open Data meaning the used URI can be universally disambiguated and machine- or human-processed. Industry approaches would typically be to use, in place of the entity reference via URI used by semantic annotation, a plain text string or an identifier from a controlled vocabulary which in term is identified by a plain text string. In such cases, it would be a standalone task to align those strings to URIs in a semantic Knowledge Base, usually with the help of a Named Entity Recognition and Linking (NER/NEL) tool. Where a controlled vocabulary is in use, this alignment is only needed to be once. A controlled vocabulary could be transformed into a Knowledge Graph and additional relations (e.g. taxonomic) extracted and represented between vocabulary items. Such a graph could also be serialised by a structured metadata model such as SKOS or RDF and stored in a Graph Database or Semantic Triple Store. Another aspect of NEL is the equivalency between entities in different Knowledge Graphs, so that the entities in the organizational graph are aligned with those in Linked Open Data graphs like DBPedia or WikiData through owl:sameAs statements (this is also the approach within ReTV where we build our own Semantic Knowledge Base for entities and capture equivalency of our entities with entities in public knowledge graphs for interoperability with other annotations).

For visual concept labels, we have noted that visual classifiers (software trained to identify certain classes of visual concept in images/videos) tend towards use of their own internal representations of each class, combined with a label (string). Online services, whether from IBM Watson, Google or others, each have their own class vocabulary with labels. We follow a concept list which has been used by various classifiers in research by different international groups, participating in the annual TRECVID evaluation workshops, thus forming effectively a standard for this area (significant is that there are also training sets for the TRECVID concepts, which act as a definition of what the visual concept actually means, whereas how IBM, Google etc. define some of their concepts is not transparent to the end user). For interoperability, one could consider aligning the concepts - 345 from the TRECVID list and 205 from the Places ontology - to Linked Open Data e.g. TRECVID concept with id 468 (label: “Taxi_Cab”) is the same as the entity Taxicab in DBPedia.

```
trecvid#468      owl:sameAs       dbp:Taxicab
```

For content categories, we have similarly noted that whereas top level Genres would be too narrow, the entire entity space of a public Knowledge Graph would be too broad, and identified the EBU Core categories as offering the right level of granularity as well as forming already an industry standard in the categorization of media content. Once again, assuming the need to interoperate between this categorization scheme and other, we could consider aligning the concepts to a common vocabulary such as a Linked Open Data dataset. Here the interpretation would likely be to map categories onto (WikiData) classes or (DBPedia) categories, i.e. considering EBU Core [2.3.12] Football (Soccer) rather than stating equivalence to the entity (instance) for the sport itself, we can map to the Wikidata or DBPedia category Association_football.

```
ebu#2.3.12       owl:sameAs       dbp:Category:Association_football
```

2.4 Results

Regarding our annotation model, we have shown in the previous section that it is feasible to map most of the descriptions to and from other standards/specifications used in the Web/TV domains. Our model combines uniquely the content description, technical metadata and program information with individual publication events (broadcast, but also Web - TVoD, archive or social media) since each publication will be associated with its own audience or success metrics (see Deliverable 2.1).

Regarding vocabularies used for metadata property values, we prefer to use entity URIs where appropriate drawn from our own Knowledge Graph (see Deliverable 1.1). The entities in the KG are linked using the owl:sameAs property to equivalents in the DBPedia and/or WikiData knowledge graphs, so that broader reasoning over the concept space is feasible. We capture both visual concepts and the content categories using controlled vocabularies which are 'standards' in their own communities (TRECVID for visual concepts, EBU for content categories) and we have noted that these vocabularies could also be aligned to entities in the same knowledge graphs if necessary (this will depend on the future requirements of the tasks on content re-purposing and recommendation - if the existing vocabularies prove to be expressive enough for their implementation).

3 Viewer Profiling

3.1 Problem Statement

We define a privacy-preserving, GDPR- and ethics-compliant viewer profile model with topical interests/preferences, identifying viewers across vectors using different identifying features (Zattoo user accounts, social logins, cookies, IP addresses etc.). We are also developing a component to generate/update viewer profiles over time by matching viewer activity (Web page comments, social media posts, TV viewing, second screen usage etc.) to interest in program topics derived from our TV program annotations. An important aspect of building viewer profiles is to minimize the manual input requirements and use a wide variety of existing sources, e.g. social media and viewer behaviour in order to populate profiles. In addition to the individual viewer profiles, this task also looks into modelling the user cohorts within each vector/channel.

3.2 State-of-the-Art Survey

Viewer profiling has been discussed in various research contexts, such as online ad targeting, recommender systems, fraud detection, social studies or health care, to name just a few. In the ReTV context, we are primarily interested in profiling online users of social media platforms and OTT (over the top) TV content.

(Dehghani, 2016) stresses the importance of the cold start problem and resulting sparsity of user profiles. The authors propose a hybrid approach of combining individual user profiles with group/segment level profiles in order to alleviate sparsity impact on the quality of content recommendation. The ReTV approach also requires to exploit both individual user data as well as segment level data (due to legal and technical constraints, as discussed in the following sections). An important conclusion from the paper is that there is an inherent trade-off between the size/number of segments and the level of personalization.

(Lin, 2015) notices the growing importance of microblogging platforms as the social counterpart of TV broadcasts. The authors propose microblogging-assisted profiling framework, that combines content profiling and user profiling in a social-aware manner (social relations between viewers as well as content propagation in the social network). Such a hybrid approach optimizing user, content and social metrics is also one of the goals of ReTV.

(Rahman, 2018) study the problem of online viewer profiling primarily in the context of optimal allocation of online resources. However, it contains valuable insights into how various attributes describing users can correlate (e.g. temporal viewing patterns with age and gender). It also proposes behavioural segmentation that can be used to drive different strategies of content summarization (task T3.3). Namely, the authors propose four segments of users: early leaving, steady watchers, highlighters and surfing watchers, who require different summarization strategies (e.g. in terms of topic variety and segments length).

Other authors (Amroun, 2017) stress the importance of profiling the user level of attention, which is especially important in the era of video consumption on mobile devices, and in the context of one of the main ReTV goals, which is video repurposing and summarization. The authors used Deep Neural Network classifier approach to identify various user behaviours before, during and after the viewing phase, also differentiating between various types of TV-enabled devices.

In the context of the optimization & recommendation of the relevant content, that will be the main goal of tasks T3.3 and T3.4, an important aspect is whether we base user profiles on explicit or implicit feedback in order to optimize relevance metrics. Since collecting explicit feedback is usually expensive and not always possible for large-scale systems aiming to optimize for multiple vectors (including social media), ReTV approach will be primarily based on optimizing implicit metrics (cf. WP2). However, hybrid approaches combining implicit and explicit user feedback are also possible (Kurapati, 2001).

3.3 ReTV Approach

There are two crucial aspects of a ReTV approach to the problem of viewer profiling:

- we need to be able to match viewer profiles to content, i.e. both viewers and content need to be described in the same attribute space
- we need to leverage implicit information on viewer interests as much as possible, thus limiting the importance of explicitly provided information, that is more expensive and time-consuming to collect (e.g. manually entered categories)

The viewer profiles will be part of the metadata model (cf. section 2.2.4) and it will use its fields. Available types of data include:

- Top-3 content categories with which user interacted in the various time horizons (last week, last month). Content categorization is created automatically, based on:
  - pre-trained word embeddings model applied to the textual transcripts (for some of TV channels, such as e.g. Swiss public channels, transcript subtitles are available directly, for some other channels they need to be created via Speech-To-Text transcriptions)
  - visual features annotation (WP1)
  - semantic augmentation of textual and visual features using semantic Knowledge Bases such as DBPedia (cf. section). For instance, by using a named entities from textual transcripts (esp. people and common names), we are able create additional attributes describing users and a content, e.g. by recognizing celebrities and discerning them from politicians.
- Viewing duration per day (overall and time-based patterns, such as weekends vs. working days)
- Favourite actors (require content annotation, cf. WP1)
- Favourite event categories (require event annotation, cf. WP2)
- Favourite video tags (require content annotation, cf. WP1)
- Socio-demographic information (extracted from aggregated audience data)
- Spatio-temporal data, also extracted from audience data:
  - when do viewers watch the content (including technical data on devices used)
  - where they watch (anonymized, segment-level geo locations)
- Social media data (statistics describing how popular a given piece of content or a topical category was across to which a given user belongs)
- Trends based on the above-listed attributes (increasing or decreasing interests)

Again, it needs to be stressed that we need to be able to match viewer profiles to content annotations so they both need to be described in the same attribute space.

We are considering a supervised and an unsupervised approach to the problem of content annotation. An unsupervised categorization approach is described in the next section. A supervised approach (training a classifier on a set of videos assigned to a topical class, or labelled with a set of tags) will be tested in the following months. Such a classifier can be based on one of the publicly available datasets, such as Youtube 8M (https://research.google.com/youtube8m)

Since the modelling phase will exploit models that are able to implicitly perform feature selection and feature engineering (such as deep neural networks or ensembles of DNNs with boosted classifiers), the plan is to collect as many attributes as it is possible with respect to the available data sources, and allow the models to automatically decide which features are important.

Collected data will be available at two different granularity levels: they either describe segments of users or individual users that view content that is analyzed and published by the TVP.

Segment-level data:

- describes a group of users, and it is not possible to identify an individual user within a group
- we will use this kind of data to describe publication vectors, like Twitter accounts, based on their similarity to a segment profile
- it is a default for the cases when we don’t have identifiers for users (due to the legal or technical constraints)

User-level data:

- describes one individual user
- possible on Zattoo platform data.
- will link user level to segment level (user may belong to one or more segments, and ideally we can link individual profiles to viewer segments)

The viewer profiles should take the richer, daily session data from Zattoo into account.

3.4 Implementation details and use

The goal of the early prototype (Fig. 2) that we created was to generate personalised summaries of video content based on viewer profiles. For example, users can generate a video each week that contains the news segments that are most interesting to them and add a couple of sport highlights. One can think of it as Spotify's "Discover Weekly" but for TV content.

The prototype implementation consists of the following modules:

- **Segmentation:** In the segmentation step we take a complete TV show as an input (one video file) and cut it into segments (multiple video files). The definition of a segment is not completely clear-cut and might vary between shows. In general, we do define a segment as a sequence of scenes that belong together semantically. For the main news show, a segment might consist of the scenes where the presenter introduces the topic, and then the actual video on the topic.

- **Tagging:** Once we have segmented the video clips, we need to understand what is happening in them. We are currently doing this using both speech-to-text and object detection. As a first step, we need to be able to detect relatively coarse categories like: "News", "Sports", "Science". In a second step, we will introduce more fine-grained categories of our own. Like for example: "Renewable energy", "AI and the future of work". This switch from broad, pre-defined categories to finer, personalised categories will happen gradually. Cf. this example[15]. We built the current categorization on the basis of EBUCore[16] genres (cf. also section 2.2.1). The approach uses an unsupervised categorization model based on multi-lingual word embeddings models provided by

[15] https://drive.google.com/open?id=10q-jXeQWTYJCJHyQvt28t3j7YVHenTtg

[16] https://tech.ebu.ch/MetadataEbuCore

Facebook Research (MUSE - Multilingual Unsupervised and Supervised Embeddings[17]). It should be stressed that a single content segment can belong to multiple topical categories (with possibly different relevance scores/weights).

- **Profiling:** Once we have the tags, we need to create a user profile that tells us which tags are interesting to which user. We can do this by explicitly asking users what topics are interesting to them (explicit mode) or by learning it from their behaviour (implicit mode)

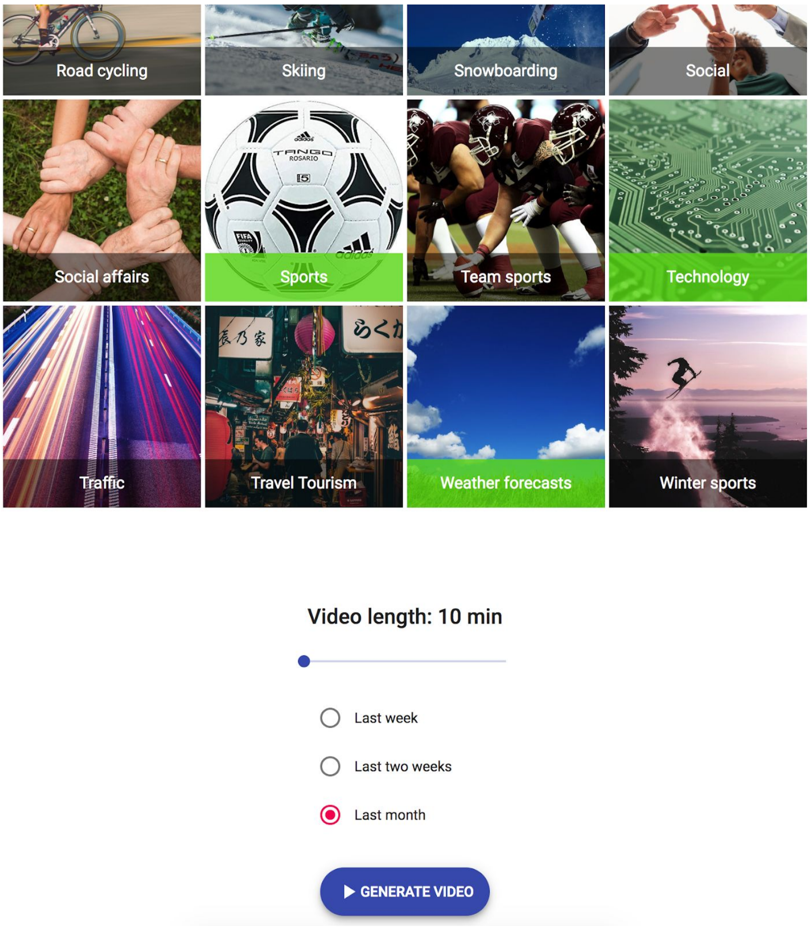
- **Personalized summarization of the content:** taking into account user-defined constraints (time horizon and expected summarization length) and available content segments (assigned to one or more topics), we have a classical optimization problem (namely, knapsack optimization) to solve. In addition to the selection of the most relevant content, we may take other metrics into account, such as topic diversification (we don't want the summary to be dominated by one topic only, even if it is the most relevant one) or serendipity (which to some extent is also related to the content recency - especially for topics like sport, it is usually preferable to include most recent segments in the summary).

The very final step is to order the selected pieces into the final personalized summary. We tested two approaches so far: grouping the segments into the coherent topical blocks, where individual pieces within each block are sorted chronologically, or alternatively, sort segments by diminishing relevance score and interleave various topics. In order to come up with the optimal approach, we will do end-user tests (also in coherence with end-user engagement monitoring, which is the use case implemented in WP5) and report the results in deliverable D3.2 at M20.

3.5 Results

The early prototype is based on explicit user feedback only (i.e., a set of user-selected categories) and was applied to the segments of one show: main news on Swiss public SRF-1 channel (Tagesschau). We use the subtitles to assign the segments to a category. One segment can be in multiple categories, and we also store how strongly we believe a segment belongs in each category. Once a user has selected a couple of categories and a length, we try to build the most relevant video out of the segments.

[17] https://github.com/facebookresearch/MUSE



Video length: 10 min

○ Last week

○ Last two weeks

● Last month

Figure 2. UI of the prototype recommender using viewer profiles.

There is obviously room for improvement. In particular, in the next phase, we would like to enable users to give feedback on how much they enjoyed the summary, so we can improve the recommendation algorithm. by tracking the viewing behaviour. We can also take implicit feedback (e.g. does the user actually finish the video) into account when re-training models. The currently implemented prototype needs to be connected with the data already available in the TVP (cf. WP4) in due course to expand the available data items for recommendation.

4 Conclusion and Outlook

Regarding metadata and vocabulary interoperability, we have provided mappings between our metadata model and others used in the Web and TV domains. We also discussed how Linked Data could be used to provide a shared machine understanding of the vocabularies we adopted for visual concepts and content categories.

Regarding viewer profiles, in the discussions we identified the following issues as crucial for the success of the task:

- we need to make sure that we can map viewer profiles to content, i.e. that we collect data that can be used to describe individual users, user segments, as well as the content pieces, content segments and content sets.

- One of our next steps should be research on the state of the art on how privacy preserving user segments can be created on the basis of aggregated social media data (such as content popularity metrics) and how they can be used to improve content re-purposing, personalized recommendations and optimal scheduling across vectors.

Finally, it should be noted that we started to work on tasks T3.3 (Content adaptation & Re-purposing) and T3.4 (Content recommendation & scheduling). The prototype described in the two preceding sections is the first step in the direction of personalized summarization (T3.3) and recommendation (T3.4). More details will be reported in the next deliverable document (D3.2).

References

(Amroun, 2017) Hamdi Amroun, M'hamed Hamy Temkit, Mehdi Ammi. Study of the viewers' TV-watching behaviors before, during and after watching a TV program using iot network. 2017 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2017.

(Dehghani, 2016) Mostafa Dehghani, Hosein Azarbonyad, Jaap Kamps, Maarten Marx. Generalized Group Profiling for Content Customization. published in proceedings of ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR'16), 2016. https://arxiv.org/abs/1609.00511

(Kurapati, 2001) Kaushal Kurapati, Srinivas Gutta, David Schaffer, Jacquelyn Martino, John Zimmerman. A Multi-Agent TV Recommender. Eighth International Conference on User Modeling: Workshop on Personalization in Future TV, Sonthofen, Germany, 2001. http://people.stern.nyu.edu/ksk227/um_2001.pdf

(Lin, 2015) Xiahong Lin, Zhi Wang, Lifeng Sun. MAP: Microblogging Assisted Profiling of TV Shows. MMM (1) 2015: pp. 442-453, 2015. https://arxiv.org/abs/1502.03190

(Markatopoulou, 2016) F. Markatopoulou et al. ITI-CERTH participation in TRECVID 2016. Published at https://www-nlpir.nist.gov/projects/tvpubs/tv16.papers/iti-certh.pdf last accessed 10 October 2018.

(Nixon, 2018) L. Nixon. Assessing the usefulness of online image annotation services for destination image measurement. In ENTER e-Tourism conference, 2018.

(Rahman, 2018) Sabidur Rahman, Hyunsu Mun, Hyongjin Lee, Youngseok Lee, Massimo Tornatore, and Biswanath Mukherjee. Insights from Analysis of Video Streaming Data to Improve Resource Management. IEEE CloudNet 2018, 2018. https://arxiv.org/abs/1806.08516

(Zhou, 2014) B. Zhou, A. Lapedriza, and J. et al. Xiao. Learning deep features for scene recognition using places database. In Advances in neural information processing systems, pages 487–495, 2014.