P2019-09 OpenLABEL

Project Proposal

Version 1.0, 07.12.2020

Table of Contents

Project Number	P2019-09
Relevant Standard	none
Project Name	OpenLABEL Concept
Project Type	Concept Project
Start Date	01.04.2020
End Date	30.09.2020
ASAM Funds	0€

Executive Summary

The OpenLABEL Concept Project will come up with a concept on how a future labeling standard could look like. This concept will include topics on labeling methodology, labeling structure, file format. The OpenLABEL Concept project will be closly coupled with tne upcoming ontology project, as the object descriptions will be delivered by the ontology project. on the other hand the OpenLABEL Project will deliver requirements to the ontology project.

Next to the Object labeling the OpenLABEL project will also cover the scenario labeling, in this case the coupling will be with the ontology project and the upcoming OpenSCENARIO project.

In the end the concept paper will cover:

- labeling methodology
- labeling structure (including relations)
- fileformat and structure definitions
- scenario labeling
  - labels derived from object labels
  - abstract labeling of scenarios

1. Overview

1.1. Motivation

This is a proposal for the development of OpenLABEL, a new standard regarding the Labeling, for Machine Learning models training and validation, of raw data generated by vehicles equipped with sensors with the capacity to enable any SAE level of autonomy >= 2.

From working with different customers, a significant fragmentation emerged in the way each individual organization categorizes and describes the objects populating the driving environment. Such categorizations and descriptions are the fundamental building block of any Autonomous Driving System's (ADS) perception stack, since it is through them that an ADS come to a primal understanding of the status of around itself, including the entities present and some aspects of their behavior. Many vital driving decisions are based on this understanding.

The lack of a common Labeling standard in the industry is the root cause of several different issues:

- **Hampered Vehicle2X Interaction**: the different descriptions/understandings of surroundings may cause casualties in complex situations involving two or more different ADSs OpenLABEL could support filling the existing V2X standards like ITS G5.
- **Precluded sharing**: It results highly difficult if not impossible to share data across organizations adopting different Labeling taxonomies and specifications
- **Lowered Annotation quality**: Each individual labeling task requires ad-hoc training and even custom software features development to be completed, that translates into a higher probability of errors and thus a threat to safety
- **Deprecation of old labels**: Long-term operation of ADS development imply changes in quantity and richness of labels to be produced, considering the evolution of the driving scenes, new sensors, and scenarios. As a consequence, a flexible descriptive language is required to absorb future extensions/modifications of labels and guarantee back-compatibility.

In sum, the absence of a labeling standard such as OpenLABEL is ultimately a significant safety threat for all road users surrounding any kind of vehicle which is being operated in autonomous or semi-autonomous (SAE Level >=2) mode. OpenLABEL objective is to increase overall operational safety by providing a language that allows for the encoding of a common baseline understanding of the driving environment for any ADS.

OpenLABEL project outcomes will include: a list of classes of interest - Labeling Taxonomy -, the structure underlying the relations among classes, the definitions for each one of the classes of interest together with examples or images and plausible class attributes, and finally the labeling specifications: a set of instructions detailing the way each class should be labeled with respect to each type of annotation, including explicit directives to treat particular instances of critical labeling situations (occlusions, associations, etc.). OpenLABEL will include the designation of a suitable data format that allows for an effective representation, storing and exchanging of the generated labels.

The OpenLABEL Concept paper will contain concepts for:

1. Labeling/Annotation Format: Specifications of a suitable data format for effectively representing, storing and exchanging the labels
2. Labeling methods: detailing the labeling approach according to the different labeling tasks (semantic segmentation, bouding box, scenario labeling…)
3. Labeling Specifications: Set of instructions detailing how to label each class with respect to the annotation task and how to treat critical labeling situations

1.2. Use Cases

Use cases in the context of ASAM standards describe the external behavior of the standardized system, i.e. the interaction of the system with a user or with another system. The description of use cases is particularly useful for explaining the motivation for new standards, major version development projects or the addition of new features in minor version development projects.

ASAM subdivides use cases into three tiers, where each lower level is a refinement of its immediate higher level. * Business usecase: Describes an economic advantage, company need, process, method or element of a larger tool chain that involves many people of a company or multiple companies in a customersupplier relationship. Example: ECU calibration and measurement. * End User Use Case: Describes a need, process, method or element of a tool chain that is handled by one person while he carries out specific tasks within a company usecase. Example: Start measuring data on an ECU. * Technical usecase: Describes a technical necessity, that is required for the operation and interoperability of technical systems, such as tools, test systems or application software, to support the tasks of enduser use cases.

This can be expanded on during a project’s development.

1.2.1. Technical use cases

Type	Technical Use-Case
Title	ML model Benchmarking
Description	Ultimately, labelled data are mostly used for the purpose of ML model Training and Validation. A unified labeling standard can allow to train and validate models on a shared description of world entities and benchmark them on a common baseline.
Actors	- AV perception developers - regulators - policy makers - local and national authorities - insurances
Notes	n.a.

Type	Technical Use-Case
Title	Machine Learning Model Validation / Ground truth for model validation
Description	To train a machine learning model, lots of labeled data is neccessary. After the training the model must be validated, for this purpose new data including the ground truth is neccessary. With standardaised labels external/new data can be used to validate the model or improve the validation and training.
Actors	- ML Engineers - Validation Engineers - local and national authorities - homologation institutions
Notes	n.a.

Type	Technical Use Case

Title	Sensor Data Annotation Review
Description	A standardized format for sensor data and sensor data annotations would help developing standard tools for visualization of the data and reviewing the annotations. This would also permit outsourcing of certain tasks within the training and test dataset creation pipeline for autonomous driving. annotate data from different sensor types (e.g. Lidar, Camera, Rader, Ultrasonic), consider time behavior of different sensor types.
Actors	- Environment Perception Developers - Data Managers
Notes	contributed by Volker Schomerus @ VW

Type	End User Use Case
Title	Reduce human learing effort for Sensor Data Annotation
Description	With a standardized format and standard sets of classes for sensor data annotation, the ordering process for sensor data annotation services would be easier and the risk of different understandings of annotation specifications could be reduced. With this the time a labeler needs to learn how the labeling works will be reduced. * enable higher quality labeling * common annotation structure * everyone follows only one guideline
Actors	- Environment Perception Developers - Data Managers - Annotation Service Providers, - Tool Developers - Labeler
Notes	contributed by Volker Schomerus @ VW

Type	End User Use Case
Title	Reduce tool implementation effort for Sensor Data Annotation
Description	With a standardized format and standard sets of classes for sensor data annotation, the ordering process for sensor data annotation services would be easier and the risk of different understandings of annotation specifications could be reduced. This has also an impact on the tool providers developing annotation tools for one labeling standards instead of having to support many different labeling structures and convetions. This will reduce costs for the whole workflow. * enable higher quality labeling * common annotation structure * everyone follows only one guideline

Actors	- Environment Perception Developers - Data Managers - Annotation Service Providers, - Tool Developers - Labeler
Notes	contributed by Volker Schomerus @ VW

Type	Technical Use-Case
Title	Metadata labelling, Multi-sensor labeling of objects and actions
Description	Objects and actions of a scene/frame need to be defined as entities, with intrinsic properties (e.g. type, name, numerical properties), and in addition, with projected features that define how are they projected/measured from different sensing devices (e.g. cameras, lasers, gps, etc.). I want to be able to label data with additional information about the data, e.g. Data Owner, Sensor type & version, Labelling algorithm version. Ideally, metadata labelling would be extensible so that I can add whatever labels I need but still be query-able in the exactly the same way that I would query on non-metadata labels. The labeling data format needs to manage such levels of information (intrinsic, measured), and allocate descriptions of the timestamps and synchronization between sensing devices, in a single payload. As a consequence, labels can be produced at different levels, by different teams, and in different time periods, for instance aggregating content from newly labeled streams as they are produced. * add metadata fields in the anntoations, the keys and field can be defined in the project (e.g. define schemas)
Actors	- Environment Perception Developers - Data Managers - Annotation Service Providers - useful for the whole workflow
Notes	merged use case contributed by Mike Freeman @ Warwick and Marcos Nieto (mnieto@vicomtech.org)

Type	Technical Use-Case
Title	Multi-type labeling of objects and semantic properties

Description	Different ADS development use cases require different type of labels to be produced (e.g. pixel-wise for semantic segmentation, 3D polygons for lane sensing, cuboids for obstacle detection, 2D polylines for pedestrian analysis, etc.). Geometric entities labeling require as well a data format that enable nested properties (e.g. visibility level, ids, confidence values, tokens). Last but not least, a single label payload of a scene may contain objective data describing objects, but also semantic concepts related to actions carried out by objects, events triggering sub-scenes, or relations between objects. The semantic level of the scene requires the existence of a governing ontology such that labels can point to concepts for further semantic consumption of the annotation files.
Actors	Environment Perception Developers, Data Managers, Annotation Service Providers.
Notes	label: - Objects - environmental conditions - Actions * - Events - relations this has to be aligned with the ontology Contributed by Marcos Nieto (mnieto@vicomtech.org)

1.2.2. Business use cases

Type	Business Use Case
Title	Dataset Sharing
Description	Sharing labeled datasets effectively and in a way that guarantees their utility across different organizations can only be achieved when such datasets are annotated following a set of standardized labels there are many datasets are available, it will enable to industry and others to easier share and use available data sets. A standardized format (classes, data structures etc.) for labels for sensor data (2D/3D objects, semantic segmentation etc.) would also allow **extending** training and test datasets for environment perception for autonomous driving.
Actors	- Academia - industry
Notes	merged with use case from Volker Schomerus @ VW

Type	Business Use-Case
Title	Tool development at Labeltool-provider

Description	Currently labeling projects differ greatly in terms of their requirements (e.g. how many label values can a single pixel have assigned for segmentations). Often tool derivatives or new features need to be implemented. This can cause bugs, increases the development time and often contradict the requirement of other labeling project requirements. By using the standard, (1) cost savings for projects following the standard are generated, because the tools can get optimized features for the Standard. (2) automated quality could get implemented. Currently approaches for automated QA are ressource (engineering) heavy and differ from project to project.
Actors	Labeltool-provider
Notes	merged with the recuding effert on the tool proivder side (1. annotation, 2. quality assurance, 3. review of the labels, 4.project definition) Contributed by Tim Rädsch @ understand.ai

Type	Business Use-Case
Title	Scenario labels to efficiently search scenario catalogues
Description	Current scenario catalogues contain scenarios at different abstraction levels with different state of completeness and target use cases. There is no standard labeling mechanism to annotate such scenarios with unique and unambiguous labels, so that the user can quickyl search scenario catalogues for a specific scenario with specific characteristics. By using the standard, (1) scenario catalogues from different companies, working groups and people can be searched by using the same labels and are therefore reusable/exchangeable (2) new scenarios can be directly annotated with the corresponding labels.
Actors	Scenario creators, scenarios users, function developers.
Notes	merge with the warwick usecase Contributed by Florian Bock (florian1.bock@audi.de)

Type	Business Use-Case
Title	Cascading labeling guidelines to the label provider
Description	After finalizing the labeling guidelines together with the customer, the labeler working on the project need to Understand and Apply The labeling guidelines. Usually Labeler get trained by their Team lead and technical systems, who cascade the information of the labeling guidelines and helps with answering questions and feedback. By using the standard, (1) Inconsistencies (both missing information and contradictory information) in the labeling guidelines (HOW and WHAT to annotate) will be avoided. (2) onboarding times for the labeler and the Team lead can be reduced since the standard should not change that often. (3) Understanding should be more unified, if the standard is used in multiple projects.
Actors	Label provider Team lead of Labeler Labeler

Notes	merge with reducing effort on the human side, every party in the workflow can follow the standard Contributed by Tim Rädsch @ understand.ai

Type	Business Use-Case
Title	Ordering labeled data as OEM from supplier
Description	Person has a budget for ordering labeled data. This is especially useful for the first time for this person to order labeled data. Person assigns the labeling task towards a Label provider. For the questions on HOW to annotate WHAT to annotate The new labeling standard will be used. By using the standard, (1) Inconsistencies (both missing information and contradictory information) in the labeling guidelines (HOW and WHAT to annotate) will be avoided. (2) iteration cycles to create the labeling guidelines will be reduced. (3) iteration cycles to review the quality will be reduced. The time to the delivery of the annotations is shortened, because (1), (2) and (3)
Actors	Person at company department (orders the data). Label provider.
Notes	Contributed by Tim Rädsch @ understand.ai styleguide for labeling including rules (do’s and dont’s) + examples

1.2.3. End user use cases

Type	End-User Use-Case
Title	Using labels to select scenarios to test an operational domain
Description	As a test engineer, I want to be able to identify a set of scenarios from a scenario database using a set of labels that I choose to define my operational domain that I want to test for. I want to select data at different abstraction levels, e.g. all types of roundabouts in the rain, 3 entrance roundabouts in drizzle, and also by specifying values, e.g. cars less than 1200mm high. * Scenario Labeling * create metadata labels for data * make scenarios searchable and compareable
Actors	Test Engineer using scenarios to test ADS
Notes	contributed by Mike Freeman @ Warwick

1.3. Requirements

The OpenLABEL Concept project shall work on concepts for a future OpenLABEL standard, so this standard can full fil the following requirments Also the following requirements are to consider:

Table 1. general requirements

general requirment	OpenLABEL shall describe the methodology how to label objects and scenarios, based on the defined ontology. This labeling should work on real and synthetic data.
general requirment	OpenLABEL shall define the required informations to indetify and label objects and scnearios
general requirment	OpenLABEL shall provide methods how to label objects
general requirment	The data format proivded by OpenLABEL to label objects and scenarios should be independet form the data source
general requirment	OpenLABEL shall get the object/termin/label definitions from the ontology project
general requirment	Data format and specification should account for and enable the definiton of objects, events, relations, actions, intentions, subject/predicat/objects tiplets (SPO) and other entties or properties allowing for a machine and human readable knowledge representation.

Table 2. technical requirments

technical requirment	OpenLABEL shall have the capability to store metadata and labels for data, independent of the source
technical requirment	OpenLABEL should support annotation of data from different source (extending and sharing data sets)
technical requirment	The OpenLABEL annotation format must be quick to serialize
technical requirment	OpenLABEL should provide metrics for quality assurance
technical requirment	In the dataformat of OpenLABEL it needs to be possible to use different labeling methods and assign relations (1:n) to labels objects, also haveing the possiblity to add actions and intentions as label to an object will enhance OpenLABEL

Table 3. End-user requirements

End-user requirement	The OpenLABEL format needs to be humand readable and easy to understand
End-user requirement	The OpenLABEL Userguide shall support the user in understanding OpenLABEL to reduce learning effort.
End-user requirement	OpenLABEL Documentation should help the user to measure the quality of the labeled dataset

End-user requirement	

1.4. Relations to Other Standards or Organizations

- Relation to the ASAM Ontology Project
- Relation to the ASAM OSI Project
- Relation to OpenDRIVE
- Relation to OpenSCENARIO

2. Technical Content

The goal of the proposed project is to create a concept paper for a future standard for Data Labeling that has to fulfil the use cases detailed above.

The OpenLABEL Concept project has the following content

2.1. Concept for an OpenLABEL Userguide "how to label"

In the project the different labeling methodologys will be research and the experience of the attending project members will helb to create a first draft of a guide for an OpenLABEL user guide. This guide will cover:

- application of the future OpenLABEL standard,
- explenations of the labeling methods (e.g 3D bounding box vs. semantic segmentation)
- guide on how to label objects in provided data (depending on the source)
- Terminology

The format structure will be created in close interaction with the Ontology project, to achive a interoperability between these to projects also the lableing structure sdhould be the same for objects and scenarios.

2.2. Labeling Specifications Desgin

Often, the frames coming from sensors such as cameras depicts complex situations going on in the surrounding of the vehicle. Objects appearing in a scene are seldom fully visible. Moreover, articulated objects , intersected objects or objects in unusual configurations can pose a threat to annotation quality and can be annotated in many different ways. Labeling specifications are a set of detailed instructions that guide the labeling process giving precise labeling directions on how to treat the various cases detailed above. As an example, instructions about how to label an articulated truck partially occluded by a car and so on.

- Develop general instructions for each class and attribute [at a meaningful level of granularity]
- Map cases, objects and attributes that require special ad hoc labeling specs and develop
- Include visuals, examples and any other asset that can help disambiguate and clarify the instructions
- Consider the time behavior of different sensor types for the annotation
- make the labeled data mergeable to extend exsisting datasets
- Enable Quality assesments

2.2.1. Object labeling

In the specifiaction design the group has to consider how to label objects in the OpenLABEL standard. The requirements from this workgroup will be shared with the ontology project.

2.2.2. Scenario labeling

In the specifiaction design the group has to consider how to label scnearios in the OpenLABEL standard. The requirements from this workgroup will be shared with the ontology project.

2.3. Labeling Data Format

In the OpenLABEL Concept project the group will come up with a proposal on what format to use for the OpenLABEL annotation format (e.g. json). Independet of the format the concept paper will also contain a first version of a possbile labeling structure e.g.:

```
metadata / header
    frame
    source
label
    type = object/scneario/...
    name (ontology link to description)
    relation
    label method
    geometry (e.g 3D bounding box)
    dynamic = yes/no
        action / intention
        ...
```

General requirements for the format and the structure are:

- make the labeled data mergeable (without converter in between) to extend exsisting datasets
- make Datasets comparable

- easy to understand
- humand readable

2.4. OpenLABEL Concept documentation

- The OpenLABEL Project will provide a Concept paper for the future standard covering the specifiaction of the format and the taxonomy (the OpenLABEL Project will receive input from the ontology project and provide requirements to the ontology project).
- Userguide ("how to label") / Styleguide for OpenLABEL
- Examples for the indivudual concepts

3. Project Resources

3.1. Required

A breakdown of the project into individual work packages and the corresponding effort required to complete them. Effort should be given in man-hours.

3.1.1. Effort

Table 4. Breakdown by Work Packages [WPs]

WP Number	1
Title / Description	Annotation Format and Structure,
Deliverable	Concept paper containing: - concept for a annotation format including a proposed structure capable to label scenarios and objects - The WP will consider exisiting labeling experience and format.
Effort (Man-days)	Estimated work effort to be performed by the service provider.

WP Number	2
Title / Description	Labeling methods for scenarios and objects, OpenLABEL UserGuide
Deliverable	Concept paper containing: - description of indetfied labeling methods for objects and scenarios - draft description of how to use specific methods for different usecases
Effort (Man-days)	Estimated work effort to be performed by the service provider.

WP Number	3

Title / Description	Labeling taxonomy, Taxonomy Structure (shared WP with Ontology)
Deliverable	Concept paper containing: - usage of ontologys for OpenLABEL - list of requirements for the ASAM ontology project - example list of required objects and labels
Effort (Man-days)	Estimated work effort to be performed by the service provider.

Table 5. Total effort

WP No.	Project member (man-days)	Service Provider (man-days)	Total (man-days)
1	28	0	28
2	28	0	28
3	28	0	28

3.1.2. Budget

This section details the budget required by the project to e.g. pay service providers and the funds to be provided by ASAM.

Table 6. Funds required for Service Providers

Task Description	Effort	Cost (€700 / man-day)
n.a.	n.a.	n.a.

Table 7. Funds Provided by ASAM

Amount (Euros)
n.a.

3.2. Committed

Member companies contribute resources for the project as per the following table.

Table 8. Work Effort

Company (Name, Location)	Committed Work (man-days)	Participant contact details (name, phone, email)
AKKA, Germany	10 days	n.a
ANSYS, Germany	6 days	Evren Yortucboylu
AVL List GmbH, Austria	7 days	n.a.

Company (Name, Location)	Committed Work (man-days)	Participant contact details (name, phone, email)
Connected Places Catapult, United Kindom	8 days	n.a.
Deepen.ai, United States	10 days	Nicola Croce
EFS, Germany	8 days	Joerg Sichermann
Peak Solution	10 days	Alexander Haßler
understand.ai, Germany	8 days	n.a.
VICOMTECH, Spain	8-10 days	Marcos Nieto, Oihana Otaegui
WMG, United Kingdom	10 days	Mike Freeman
Total	**86 days**	

The following intellectual property will be transferred from member companies to ASAM:

Table 9. Intellectual Property

Company (Name, Location)	IP Description	Value (Euros)
n.a.	n.a.	n.a.

3.3. Summary

Table 10. Required work effort should be less than or equal to committed work effort + service provider contracts

Committed Work Effort	86 days
Contracted to Service Providers	0 days
Required Work Effort	84 days

4. Project Plan

4.1. Timeline

The work packages shall be carried out as per the following time schedule:
All the WP will run in parallel till the end of the project.

Month	April	May	June	July	August	September
WP 1.	X	X	X	X	X	X
WP 2.	X	X	X	X	X	X
WP 3.	X	X	X	X	X	X

4.2. Deliverables

At the end of the project, the project group will hand over the following deliverables to ASAM:

Item No.	Description
1.	Concept paper for OpenLABEL
2.	List of requirements for the ASAM Ontology project

4.3. Review Process

The following quality assurance measures shall be carried out by the project:

□ Peer Review

□ Editorial Review

□ Project Internal Review

□ Public Review

□ Reference Implementation

□ Implementation Project

□ Validator Project

□ <Other QA measure>