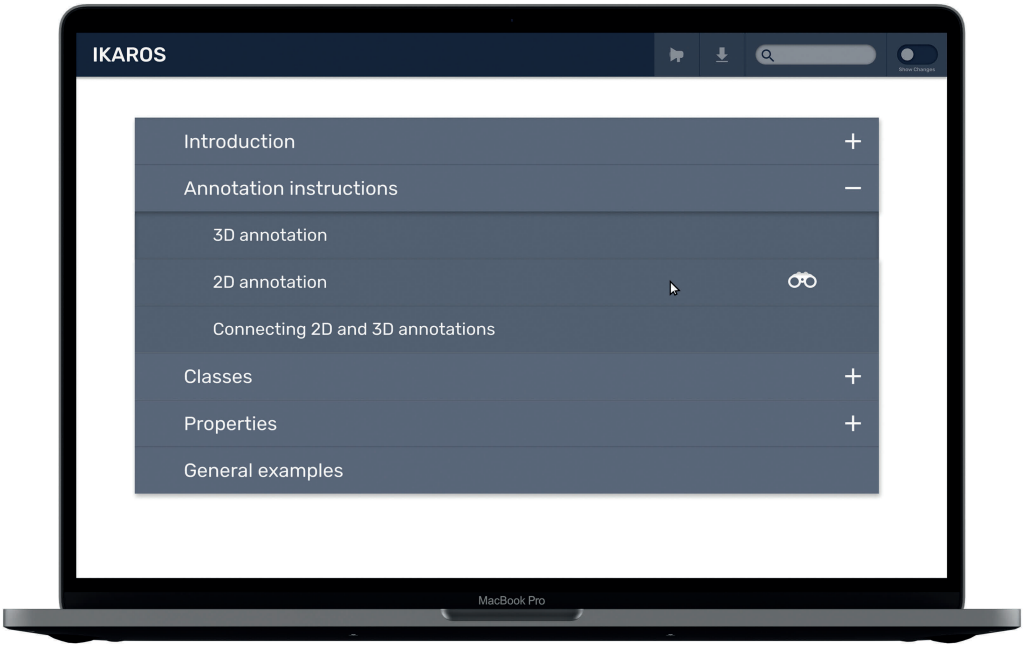
CHALMERS
UNIVERSITY OF TECHNOLOGY



Creating an interactive tool that accelerates the learning of annotation instructions

Master's thesis in Industrial and Material Science

PONTUS HOLMGREN
MADELEINE ÅSBLOM

DEPARTMENT OF INDUSTRIAL AND MATERIAL SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

MASTER THESIS 2022

Creating an interactive tool that accelerates the learning of annotation instructions

PONTUS HOLMGREN

MADELEINE ÅSBLOM

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Industrial and Material Science
Division of Design & Human factors
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022

Creating an interactive tool that accelerates the learning of annotation instructions
PONTUS HOLMGREN, MADELEINE ÅSBLOM

Supervisor: Oskar Rexfelt, Department of Industrial and Material Science
Examiner: Oskar Rexfelt, Department of Industrial and Material Science

Master's Thesis 2022
Department of Industrial and Material Science
Division of Design & Human Factors
Chalmers University of Technology
SE-412 96 Gothenbuth
Telephone +46 31 772 1000

Printed by Chalmers Digital Printing
Gothenburg, Sweden 2022

Creating an interactive tool that accelerates the learning of annotation instructions

PONTUS HOLMGREN, MADELEINE ÅSBLOM
Department of Industrial and Material Science
Chalmers University of Technology

Abstract

Machine learning is a rapidly growing industry and plays a crucial role to make autonomous mobility safe and possible. To create these machine learning models, high-quality annotations are of great importance. One company that provides this type of data is Omara. The annotation process is a human-driven process where data is labelled based on objects of interest for that specific machine learning model. For the annotators to know what and how they should annotate the data, instructions are required. The communication of these instructions, which today is via PDF files, is the area of interest in this project.

The goal of this project was to suggest how instructions can be communicated in order to accelerate the annotators' learning process, and thereby potentially increase the quality of the annotations and reduce the time needed to produce them. This was conducted by initially executing a user study where the current situation was analysed, and problems identified. This led to the creation of a list of requirements which was the foundation for the ideation and creation of concepts for a future solution. Through theoretical- and user evaluations all but one concept could be filtered out, a web-based UI that facilitates the instructions needed to annotate. The UI helps to meet requirements by, to name a few of the most important aspects, providing a consistent user experience, an easy overview of the instructions, and possibilities to easily update. This concept was then refined, and a high-fidelity prototype in Figma was created.

Keywords: Annotation instructions, Communicating Instructions, User Interface (UI), User Research, Prototyping

Acknowledgements

This master thesis was written at the Department of Industrial and Material Science at Chalmers University of Technology, by students from the master programs Product Development and Interaction Design and Technologies.

We would like to thank our company supervisors at Omara, who introduced us to the project and provided us with guidance through every step. We would also like to thank our academic supervisor and examiner Oskar Rexfelt, who supported us throughout the project. Without their knowledge, support, and advice this project wouldn't been possible.

Thanks to all the people at Omara who welcomed us to the company and helped and supported us executing this project. We also want to send a thank you to our opponents, Hanna Cervak and Lovisa Sköld, who provided great feedback and comments.

Finally, we would like to thank everybody who has participated in our user studies. Without your knowledge, insights, feedback, and opinions this project wouldn't have been possible!

Gothenburg, June 2022

Table Of Content

Vocabulary & Abbreviations

Term (Abbreviations)	Explanation
Annotation	The processes of assigning data to pictures/video
Annotator	Creates the annotations
Business Process Outsourcing Quality Managers (BPOQM)	Business process outsourcing. Responsible for the annotators delivering according to quality and time requirements. They annotate, corrects annotations, provides feedback, and coaches the annotators. They are the ones in contact with the IQMs for discussions.
Delivery team	An Omara team, responsible for the delivery organisation i.e. annotators, QMs and IQMs. They asses how big of a team that is needed, who should be part of the team, if they need training, when new people should be added to the team, etc.
Golden annotations	Correct annotation used to help calculate quality statistics.
Ground truth data	The annotated data that contains the objective truth of the state of objects, based on a human's classification of the objects.
Inhouse Quality Managers (IQMs)	Omara's expert annotators. They know the tool, should understand the guideline, and help BPOQMs understand what to annotate and with what quality and speed. They correct examinations, make golden annotations, evaluate guidelines and requirements, etc.
LiDAR	Light Detection and Ranging, a remote sensing method that uses light in the form of a pulsed laser to measure ranges. It generates a cloud of points in 3D where each point in the point cloud is one detection of a surface in the real world.
Machine Learning (ML)	The use of data algorithms to mimic the way which human learns.

Inhouse Expert (IE)	Works with annotated data and ML models. Takes part in defining the guideline together with customers.
Taxonomy	A translation from the guideline to the annotation platform. The tasks in the platform (for example, which classes exists, properties etc.) are defined based on the instructions in the guideline.
Training	A process prior to the production, where the annotators has received the GL and use it to train on annotating according to the GL instructions. The training process is summed up in a test which the annotators must pass in order to start the project.
User Documentation	A website containing complementing information to the user to help understand the tool, how it works and how it should be used effectively in different situations

01.

INTRODUCTION

In this chapter, an introduction to this project will be presented. The chapter aims to put the project into context by presenting its background and the problems. Furthermore, a description of the underlying process, which this project aims to facilitate, will be presented followed by the goals of the project. Delimitations and the process which this project has followed will also be explained.

START

1.1 Background

Data annotations are a fast-growing market. According to a recent report the global market for annotation tools is estimated to increase from 1 billion USD in 2021 to 10 billion USD by the year 2028 (Wadhwani & Loomba, 2022). Annotations are a vital part of machine learning (ML) and are needed to be able to create ML models that safely and reliably can interpret data input. The annotation process is a partly manual process where annotators mark, classify and give properties to objects in images, videos, and 3D point clouds. This data, the ground truth data, is then used in the ML model.

There exist several companies that specialise in providing ground truth data to customers. Omara, which this thesis was written in collaboration with, is one of these companies. Omara's product is the data they provide to their clients which is annotated according, or as close as possible, to the client's wishes and demands. Omara provides a platform and management, the actual annotating process however, is outsourced to companies around the world.

In a feasibility study carried out prior to this project, it was found that an important part of the annotation process is the so-called guideline (GL). The GL is used in all stages of the processes and links the client's demands with the actual data produced by the annotators. It is the customer that provides the GL to Omara which is then used before, during, and after the production of the ground truth data. This means that it is of great importance that the GL can communicate the client's demands clearly to the annotators, so that the annotators can produce the data with the required speed and quality. According to Omara, one complex factor is that the clients write the guidelines themself, which means that no guideline is the other alike. There were indications showing that there might be a lack of consistency between the GLs. Depending on the project the length of the GL can also vary greatly. Furthermore, in a long PDF-document, the search for information can become cumbersome.

An early general project description was handed from Omara to the authors of this thesis where the goal was to help accelerate the learning process for the annotators. Further, during the feasibility study, the field of how to accelerate learning was narrowed down to focus on the GL. This is mainly due to the GL being central throughout the process and due to it containing the entire work descriptions for the annotators.

1.1.1 The Company and the problem

Omara is a relatively new company that was founded based on the difficulty of obtaining training data of high quality and reliability. Omara has grown rapidly and has employees, clients, and relations worldwide. With the growth and increased number of people, not only employees but also outsourced people and customers, involved in the process of producing ground truth data the GLs plays a major role. It is the one document that describes what is supposed to be done and travels from the early customer needs to controlling the data produced, which means that a lot of people are highly dependent on the GL. If something is unclear or missing in the GL, the GL needs

to be updated, alternatively instructions are required to be communicated via separate channels. This increases the risk of the information becoming unstructured and scattered, where instructions are communicated in different ways and in different channels. If the instructions are communicated in a way that limits the need for clarifying communication in separate channels, a more rapid and effective process can be obtained. In conclusion, the accuracy and layout of the GL directly affect the speed and quality of the data produced.

1.2 The annotation process

To understand the purpose and goal of this project a description of the annotation process is deemed necessary. This description will cover the basics regarding how the actual annotated data is produced from a technical viewpoint.

The annotators work in a web-based platform which is the Omara app. In this platform, the annotators can access the project they are assigned, with the corresponding data which is to be annotated. To put it simply the goal of the annotation is to mark data in an image, video frame, or a 3D-point cloud called a LiDAR view. What and how this data should be marked depends on the project and the customers' demands. For example, the task might be to annotate all vehicles and pedestrians, road signs, facial expressions, type of road/area, etc.

Since the task itself can differ majorly from project to project, the workflow of the annotation can also vary. But in general, the following steps are performed in an annotation. First, the annotator decides which object should be annotated and determines which class this object belongs to for example, "automotive" or "pedestrian". When the object has been marked and the corresponding class selected the annotator moves on to provide more information about the objected with different properties. For example, if there is a class called "automotive" a property to that class might be "type" and values for that property be "car", "van" and "truck". After all the properties have been determined, the annotator moves on to the next object until all the objects have been annotated.

1.3 Goal

The purpose of this project is to investigate and suggest how instructions can be communicated to annotators across the globe with various backgrounds, in a user-friendly way that accelerates learning. This, mainly by looking into the GL and by further analysing how it is used today as well as applying a human-centred approach for creating a new solution based on learnings made throughout the project. All with the intent to find a solution that should be beneficial for all stakeholders.

An efficient solution to the GL can help streamline the annotation process and create better conditions for the annotators' learning process of instructions. Ultimately this would shorten the time to annotate data as well as produce higher quality results of the data annotated. The knowledge that will be gained from this project has the possibility to be applied to contexts where annotations

are used in general. To some extent, the learnings from this project could be applied to a vast variety of fields where communication of instructions is used, and not only in relation to annotators and deep learning.

The questions that this project seeks to answer are:

- How is the Guideline used today and what problems exist in relation to it?
- How can a new solution be designed that increases the usability and accelerates the annotators' learning process of instructions?

1.4 Delimitations

The platform in which the annotators annotate is to be unchanged and seen as outside of the scope for this task. Additions to it may occur but fundamental changes to the platform would imply too big of a change to the product Omara provides.

Furthermore, this project will focus on the learning process of the instructions by investigating how the instructions are communicated to the annotators. This means that the creation of the actual instructions lays outside the scope of this project, mainly due to them being created by the customers to Omara. In general, the impact that a future solution might have on the customers will be kept to a minimum, since the financial and business relationship is aspects that won't be considered in this project.

1.5 Project process

The process of this project is based on the double diamond model. This model proposes a way of thinking that first focuses on divergent (exploring) thinking and then on convergent (focused) thinking (Design Council, 2015). The double diamond consists of four phases (Design Council, 2015):

- Discover: Understanding the problem.
- Define: Based on the learnings from the discovery phase define the problem.
- Develop: Find and develop different possible solutions to the problem.
- Deliver: Test the solution, refining the ones that work and rejecting those that don't.

Worth pointing out is that this isn't a linear process but an iterative one. The different phases were executed multiple times and learnings from the previous executions were incorporated into the next ones. A schematic of how the double diamond model were used in this project can be seen in Figure 1. A brief explanation of the goal of each phase, adapted to this project, will be presented below:

- Discover: In this phase the main objective was to understand how the GLs are used today and the different stakeholders involved in producing, using and maintaining the GLs.

Interviews and a survey were used to collect the data. This resulted in a vast verity of quantitative and qualitative data.

- Define: The goal of this phase was to, based on the findings in the previous phase, define the problem in a concise way. To define the problem an affinity diagram was used, which also were used as a foundation for creating a list of requirements. To define the requirements of the different stakeholders, personas were created.
- Develop: In the development phase several different concepts were created which aimed to solve the problems defined in the earlier phases. This was done by executing ideation methods which led to the creation of multiple concepts. To select which one of the produced concepts to move forward with theoretical evaluations and an evaluation workshop was carried out.
- Deliver: The final phase aimed to refine the final concept by creating a Figma prototype. This prototype was then evaluated with a usability test where strengths and weaknesses could be identified.

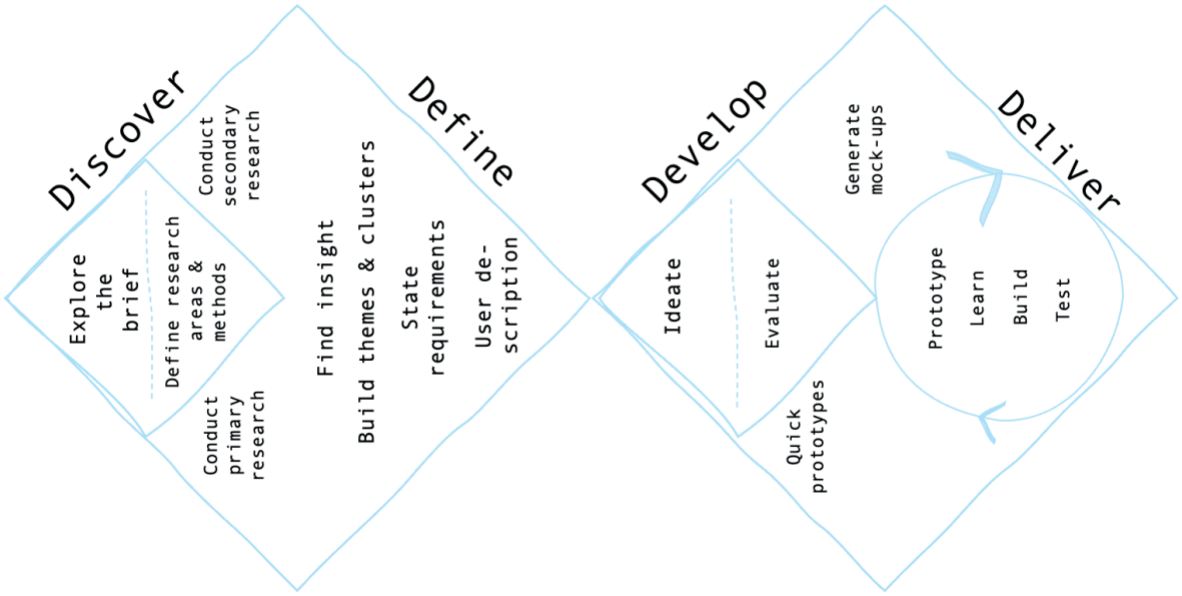


Figure 1 - Schematic of the process, based on the double diamond model.

02.
THEORETICAL FRAMEWORK

In this chapter, the theoretical framework relevant for conducting this project will be presented. It focuses on presenting how the human interacts with technologies by looking into areas such as cognitive factors, the learning process, and the usability of a product.

2.1 Human-machine system

According to the American Psychological Association (n.d.) the definition of a human-machine system is "in ergonomics, any system with interdependencies between human operators, machines, and processes" and an efficient human-machine system comes with great benefits. A human-machine system that is optimally designed should lower the chances of errors that could be devastating. If it is optimally designed it should also lower the usage time and increase competitiveness as well as making the user feel more confident, summing up to increasing their overall mental state. For a human-machine system to be optimally designed it needs to be useful, meaning that it needs to be able to perform specific tasks and achieve desired goal. Whether it is useful depends on the systems' utility and usability (Osvalder & Ulfvengren 2015). Where utility is defined as "whether it provides the features you need" and usability as "how easy & pleasant these features are to use" (Nielsen, 2012), read more about usability in 2.4 Usability.

2.2 Cognitive ergonomics

Cognition is, according to the Cambridge Dictionary (n.d.), defined as "the use of conscious mental processes" and includes mental processes like sensation, perception, and attention. Cognition however can also include more complex operations like how to solve a problem, reasoning, memory, etc. (Smith & Kelly, 2015). A cognitive process is the process of understanding or ignoring the stimuli given by the human senses. An important aspect of this is the so-called bottom-up and top-down processes. The bottom-up process refers to the process of simply taking in the stimuli which is available for the senses, without giving them any further analysis. This means that bottom-up processes don't rely on any previous knowledge or experiences, but solely on the information which is presently available by the sense. The top-down process on the other hand begins with previous knowledge and experiences. By actively interpreting and analysing the information available by the senses, based on previous knowledge, sense can be made for the individual from incomplete information (Osvalder & Ulfvengren 2015; Rousay, 2021). For example, if a human was presented with the incomplete sentence "There wasn't a cloud in the sky, only the warm ___, which shined all day long" most people would probably be able to understand that the missing word in the sentence is "sun". We, as humans, can do this by using our previous experiences and knowledge of, in this case, the description of different weather phenomena. When information is perceived usually a combination of bottom-up and top-down processing is used (Osvalder & Ulfvengren 2015).

Stimuli can be perceived by all the senses the human body can experience, but the most important one is vision since 80% of the sensory inputs come from vision (Haupt & Huber, 2008). The process of using the vision to detect certain objects can be divided into search and scanning. When scanning the information is gone through systematically with a clear understanding of what is being searched for. For example, looking for one's name in a table seating chart. Searching, on the other hand, is more random and there exists a bigger uncertainty regarding what the searched object looks like. Two important aspects related to the visual search are how obvious the searched object

is and what the person searching expects the object to look like as well as where he/she expects it to be found. In turn, this means that if the object is placed with too similar information, it would take a longer time to find it, something worth considering when designing UI in order to reduce the risk of information overload (Osvalder & Ulfvengren 2015).

Related to visual stimuli Osvalder and Ulfvengren (2015) present a set of design principles when designing visual displays, where they state that an overarching rule is to "…present the right information, in the right way, at the right time". This is a selection of the principles which was deemed most relevant for this project:

- **Minimised the time and effort required for finding information:** Information should be a group if they belong together and not be scattered in different views, displays, menus, etc. Furthermore, information that is used frequently should be easy and quick to access.
- **Proximity:** When two different sources of information need to be used, they should be placed near one another. This doesn't only aim towards physical distance, but also proximity regarding colours used, patterns, highlights, etc.
- **Knowledge of the world around us:** Generally, the user shouldn't have to keep a lot of information in the short-term memory, or use memories stored in the long-term memory, since this capacity instead can be used to solve the task at hand. Instead of relying on the user's memory, the necessary information should be presented to the user. Worth mentioning is that if too much information is presented this could create an overflow of information, making it hard to interpret. Hence, a balance needs to be maintained.
- **Anticipate system statuses:** It can be difficult for the user to predict what will happen next based on the current system status, since this requires the user to use several sources of information at the same time while anticipating how the system will function in the future based on this information. It is therefore important to help the user to be proactive by presenting possible outcomes based on the current and previous statuses.
- **Consistent presentation:** Some tasks, which are frequently executed by the user, become stored in the long-term memory and the user can start performing these actions automatically. It is therefore important to be consistent with how similar interfaces are designed. These can be aspects that are more general to most users, for example, the use of the colours red and green and symbols. Others might be more specific to that specific user type, for example, the uses of colours in a workplace-specific system, placement of emergency buttons on certain machines, etc.

2.3 The process of learning

According to Osvalder and Ulfvengren (2015) the memory can be divided into long-term and short-term memory, where the long-term memory can store memories for as long as a lifetime while the short-term memory is more of an active process that only can store and process memories for a shorter time period. Furthermore, Osvalder and Ulfvengren state that it is the information available in the short-term memory on which we humans base our decisions on. To be able to use

information, which isn't currently available via our senses but previously has been learned, two things are required: encoding and retrieving. Osvalder and Ulfvengren describe the two concepts as follows. Encoding is the process of storing a memory from the short-term memory to the long-term memory. Retrieving is when a memory in the long-term memory is being activated and brought into the short-term memory. Cues can help us to retrieve memories from the long-term memory and these cues are determined based on the sensory information at the time of storage. For example, a specific place, sound, smell, or feeling can be a cue for certain memories.

Osvalder and Ulfvengren (2015) offer guidance on how systems can be designed to support the memory process through associations and consistency. Since the long-term memory is structured like a network, by providing associations in a meaningful way a system can help the user to create strong networks. This can be done by categorising the information and by avoiding abstract descriptions instead of using concrete ones. Another strategy is to encourage the user to create summaries to make it more likely that strong associations are made in the network. Lastly, they suggest that the design should be consistent with similar systems to make it easier for the user to recognise based on previous experiences.

2.4 Usability

Nielsen (2012) defines usability as a "quality attribute that assesses how easy user interfaces are to use" and that it is based on 5 quality components: learnability, efficiency, memorability, errors, and satisfaction. He continues to argue that usability together with utility is what defines whether an interface is useful. If a system can achieve the desired task but the interface is too difficult to use the system fails. Similarly, it won't matter if the interface is easy to use if the system is unable to achieve the desired task. To achieve a useful interface Nielsen has made some guides, where one is presented below, see 2.4.1 Nielsens 10 Usability heuristics.

However, usability is also defined as "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" according to the International Standards Organisation, ISO (2018). A definition that Jordan (2002) has worked rigorously with. Also, Jordan has broken down usability into 5 components: guessability, learnability, experienced user performance (EUP), system potential, and re-usability. Jordan has also created principles that can be adapted to create usable design, see 2.4.2 Jordans principles of usable design.

2.4.1 Nielsens 10 Usability heuristics

Nielsen has produced 10 usability heuristics that can be used as guiding rules when designing interactive products. Below follows a list of these heuristics and a short explanation of each of them:

- Visibility of system status: Through relevant and appropriate feedback the design should inform the user about what's happening. This in turn means that the user should be able to learn from previous interactions with the design and be able to determine what to do next.
- Match between system and the real world: The design should stay true to real-world conventions and use language and concepts which are familiar to the user. This makes it easier for the users to learn how to use the interface since it is familiar to the user and functions in a way that the user expects it to function.
- User control and freedom: Users should be able to quickly and easily to able to quit and leave an action that the user performed by mistake. This means that it should be easy to cancel, undo, exit, etc.
- Consistency and standards: The design should follow conventions established by other similar designs and industry standards. This means that the interface more easily meets the expectations of the users regarding how they think the interface should work.
- Error prevention: The design should prevent errors from occurring. This can be done by removing certain conditions where errors can occur or by informing the user about the action and present options before they engage in this action.
- Recognition rather the recall: The design should minimise the amount of memory the users have to use during the interaction. This means that the information the user needs should be easily accessible and the need for memorised information should be limited.
- Flexibility and efficiency of use: The design should allow more experienced users to use the system more efficiently. This can be done by adding shortcuts or other hidden features.
- Aesthetic and minimalist design: The information presented in the design should be relevant and useable for the user. Every piece of information that is presented to the user competes for the user's attention.
- Help users recognise, diagnose, and recover from errors: When errors occur the error message should help the user to understand what has gone wrong and how the user can make it right. This means that the error should be described in an understandable way and suggest a solution
- Help and documentation: Most preferably the design doesn't need any extra help and explanation. But sometimes it might be required to provide instructions regarding how a task should be preform. These instructions should be easy to search for and concise.

(Nielsen, 2020)

2.4.2 Jordans principles of usable design

Like Nielsen's usability heuristics, Jordan provides 10 principles that affects the usability of a product. Below a summarised description of these principles is presented:

- Consistency: This means that the product should be designed in a way that solves similar tasks within the product in similar ways, for example, information being displayed in similar ways regardless of where in an interface one is situated.

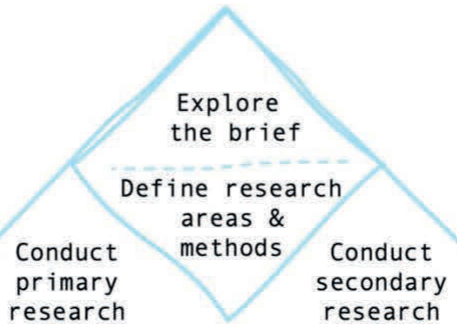
- Compatibility: Is aiming for a product to be designed to solve a problem with methods that resembles how the user expects the problem to be solved. Uses knowledge of how other products work to solve the problems, for example, the gas pedal in a car and the gas pedal for a sewing machine
- Consideration of user resources: Considers the users' resources during the interaction. For example, a voiceover system in a car where the eyes need to be focused on the road and hands fixed on the steering wheel.
- Feedback: Gives the user acknowledgement that the product has registered the operation and generates information regarding what has been registered. For example, a stop button on a bus that makes a sound and then lights a stop sign.
- Error prevention and recovery: Refers to designing a system, product, or interface to minimise the chances of wrongful usage that generates errors. Furthermore, it also refers to if an error does occur, recovery from that error should be fast and easy. An example of error prevention can be displaying the format which the social security should be inputted. An example of recovery is “undo” buttons.
- User control: This refers to a product being designed in a way in which the user feels as though he/she has control over the tasks the product is designed to fulfil. For example, the windshield wipers of a car where they can be put on auto or run manually, and the user can affect the speed as well.
- Visual clarity: Means that the product should be designed in a way that makes the displayed information readable in a fast way without disturbance and confusion. This can be exemplified in an example of choosing how to display information. A digital screen might not be optimal in direct sunlight, but a sign or sheet might not be functional if darker conditions.
- Prioritisation of functionality and information: Refers to designing a product in a way that prioritises functionalities or specific information. An example can be road signs where all information is too much to handle so a prioritisation of information is made. Where the intended person should know whether he/she is heading toward any of the bigger cities and the closer one gets the more specific information is presented.
- Appropriate transfer of technology: Means that one should make use of existing technology to enhance the usability of a product. Such as digital screens for scales/thermometers for more correct measurement and easier readability.
- Explicitness: Refers to designing a product so that clues are indicating the functionality of the operations. It can be done with different methods such as using symbols or design laws like colour coding. An example can be water taps that with symbols indicates still and fizzy water, where the shape of the tap itself indicates that the tap can be pulled, and a symbol indicates what will come out of the operation (still or fizzy water).

(Jordan, 2002)

03.

DISCOVER

In this chapter, the discovery phase will be presented. During the discovery phase the subject was explored through several methods with the ultimate goal to deliver unstructured research findings.



3.1 Method

Below the methods used in the discovery phase will be presented. These methods and their results acted as a foundation for the entire project.

3.1.1 Interviews

Interviews can be described as a meeting where the goal is to learn more about the users' motivations, opinions, attitudes, etc. (Wikberg Nilsson et al., 2015). Interviews can be divided into 3 different types: unstructured, semi-structured, and structured interviews (Preece et al., 2015). Which type of interview is most suitable depends on the purpose of the interview. If the goal is to get a first impression about a subject an unstructured interview might be more suitable, while a structured interview might be a better approach if the goal is to know more about a very specific subject (Preece et al., 2015).

An unstructured interview is the most exploratory type of interview and an interview where the interviewer has the least amount of control of the process since the questions are open-ended and the answers might be hard to predict before the interview starts (Preece et al., 2015). This means asking questions like "why?" and "can you tell me more about…", so-called probing, is an important part of unstructured interviews (Preece et al., 2015). These types of interviews might be useful at the beginning of a project when the knowledge of the subject is limited and specific questions hard to formulate.

On the other side of the spectrum, structured interviews can be found. In these types of interviews, the interviewer asks the same questions to all the interview subjects, and the questions are often closed, meaning they have a pre-determined set of answers, for example, "yes"/"no" questions (Preece et al., 2015).

Semi-structured interviews are a combination of these types of interviews, containing both open and closed questions (Preece et al., 2015). A basic script is often used to make sure that the same topics are covered in all the interviews, but probing can still be used to find out more about certain subjects and new information that might arise (Preece et al., 2015).

Table 1 - Table of interview subjects.

Subject number	Role
1	IE
2	IE
3	IQM
4	IQM
5	IQM
6	BPOQM
7	BPOQM

In this project, interviews have been used multiple times to investigate what could be improved during the flow of information and go deeper into the usage of GLs, see Table 1 for a specification of all the interview subjects. The goal of the interviews was primarily to understand who and how the GL was used during the entire annotation process, what areas of improvements the GL had, and what requirements exited for a future solution.

The type of interviews used in this project were mainly semi-structured interviews. Before the interviews were conducted, information about the situation and problem had been investigated

through a couple of informal unstructured interviews. This meant that some information about the area of interest existed and a questionnaire for semi-structured interviews could be created. However, since the information about the GL still was limited and the purpose of the interviews was to explore and investigate what problems existed, semi-structured interviews were deemed to be an appropriate approach. This ensured that the same topics were covered in all the interviews, but probing was still used to investigate previously unknown problems or subjects that arose during the interviews. The interviews were conducted in the same order as presented in Table 1.

The reason for choosing to interview the subjects in this order and start with the IEs was due to them having a wide view on the GL. This provided general information regarding the GL as well as an overview of how it is used during different stages of the annotating process. IQM and BPOQM are closer to the actual annotation production and therefore they were interviewed in a later stage when more information about the GL and the flow of information had been acquired.

To ensure that all information was perceived correctly all interviews were recorded with consent from the interviewees and later transcribed.

3.1.1.1 Pre-Recorded interviews

The aim was to conduct interviews with customers but due to time restraints of the customers this wasn't possible. However, Omara, the company in which this project was conducted in collaboration with, has previously conducted research regarding guidelines. It was found that Omara had conducted interviews with customers related to the topic of GLs, which were recorded. These interviews were transcribed as well and used in the define phase.

3.1.2 Survey

Like interviews, surveys can have open and/or closed questions, but the difference is that the survey can collect data from a high number of participants (Preece et al., 2015). One of the biggest advantages of a survey is the possibility to get answers from a large group of people without the same amount of labour needed if the data was collected via interviews. However, since there is no interviewer present when the participants answer the survey, there's a risk of misinterpretation and data might be lost since probing and follow-up questions aren't possible (Preece et al., 2015).

In this project, a survey was sent out to the annotators. The main reason for sending out a survey to the annotators instead of using interviews and/or observations was their inaccessibility. Since the annotators work in other countries than the Omara and aren't employed by them it was difficult to get access to them for an interview. The survey was sent out after the previous interviews had been conducted and information about the problems with the GL had been gained. The main goal of the survey was to collect opinions from the annotators regarding how they use the GL in their work today and what opinions and problems they have had with this solution. The survey was sent out to around 250 annotators and 228 of these responded.

To gather quantitative data that rapidly could be analysed most of the questions in the survey were closed questions and Likert scales were used in multiple of the questions. The survey began with questions related to the annotators working experience and language skills, with the purpose of understanding how long the annotators have worked for Omara and how comfortable they are with the English language. The survey then moved on with more specific questions about the GL and how satisfied the annotators are with this solution. Using a Likert scale the annotators were asked to rank how well they thought the GL is at communicating instruction, how easy it is to comprehend, how easy it is to find information, amongst other things. The annotators were then asked for possible rooms for improvements in the GL and if they could name specific projects where they found the GL to be extra good or bad. Questions were also asked regarding the annotators working procedure in general, for example how they dealt with problems and uncertainties, how they mainly received instructions, and what type of errors they find to be most common. While the ambition was to include as many closed questions as possible, some of the questions were posed as open questions due to the knowledge of certain topics being limited. Hence, a big mix of different types of questions was used.

3.2 Results

Below the results of the Discover phase will be presented. The findings are broken down to be presented as the different roles, the annotators situation, and the annotators view of the guideline

3.2.1 The different roles

One of the first findings in the project was the mapping of the stakeholders who in one way or another use the GL. This information was important to understand how the GL is used today and what requirements a future solution must fulfil.

Worth mentioning is that there exist several other roles within Omara than what is presented here, for example, customer success manager, which communicates with the customer on a less technical level, and the Delivery Team, which is responsible for the production. These roles have not been included in this description since their interaction with the GL is limited and therefore deemed outside the scope of this project.

3.2.1.1 Customer

It is the customer that creates the GL and therefore owns the document. When the customer creates the document one key aspect from the customers' point of view is to describe what and how they want the data to be annotated, for them to get the delivery they want.

3.2.1.2 Inhouse expert

The IE communicates a lot with the customers regarding the GL and can be seen as the person that has the main responsibility for the GL during the production. They are responsible for translating the GL into so-called Taxonomies, i.e., converting the guideline into tasks in the Omara app which is what the annotators see when they are annotating. For example, if the customers want the class

"automotive" to have the property "type" with the values "car", "van" and "truck", it is the IE that makes sure that the annotator can annotate the data according to these requirements.

Before the production starts the IE has a lot of communication with the customers to refine and clarify the GL. The length of this process depends on the project as well as the state which the GL is in when delivered to Omara. Sometimes the GL can be written more from the perspective of what data the customers want delivered, rather than how the data should be annotated.

> *"Sometimes they are quite…How should you put it… written from what they [the customers] expect for delivery. And sometimes they are written from instructions to this is how you should do it." (Inhouse expert, authors translation)*

This means that some GLs might take several iterations before they are ready for production. In some cases, the IE even rewrites the GL in order to make them communicate annotation instructions rather than specifying data that should be delivered.

> *"Sometimes I rewrite them myself, starting from the beginning or do some editing in theirs, or sometimes they are so good that they can be sent directly to the annotators." (Inhouse expert, authors translation)*

Throughout the production process, it's the IE that has the main responsibility for the GL. This means the IE might need to answer questions that couldn't be answered by persons further down the production process. In some cases, an update of the GL might be required, and in such cases, it is the IE that is responsible for making sure this update is made.

3.2.1.3 Inhouse quality manager

The IQMs are employed by Omara but aren't stationed at Omara's office, at the time of the interviews they were mainly stationed in Ukraine. The IQM is responsible for the BPOQMs and annotators. When a new project starts, the IQM is a part of the refining process of the GL and provides feedback to the IE regarding what needs to be clarified. The IQM makes sure that the annotators and BPOQMs understand the GL and answer questions that might arise during the production. The questions are mainly raised to the IQMs via the BPOQM or in general meetings with all the annotators. It is also the IQM that calculates statistics and provides feedback to the annotators. The number of annotators the IQMs are responsible for varies from project to project, the interviewed IQMs managed from 12-35 annotators.

3.2.1.4 Business process outsourcing quality manager

The BPOQMs are the annotator's closest manager, and they are managing the day-to-day operations. This means answering questions from the annotators and providing feedback regarding speed and quality to the annotators. The BPOQMs aren't employed by Omara, but instead by the company from which Omara buys the annotation service from.

3.2.1.5 Annotators

The annotators are the ones that produce all data that the customers have ordered. Like the BPOQMs the annotators aren't employed by Omara directly. They use the Omara app to produce the annotated data and use the GL as their main source of instructions. But they also receive instructions from the BPOQM and sometimes IQM. Before they can start producing the data they receive training on the GL and is required to pass an examination test.

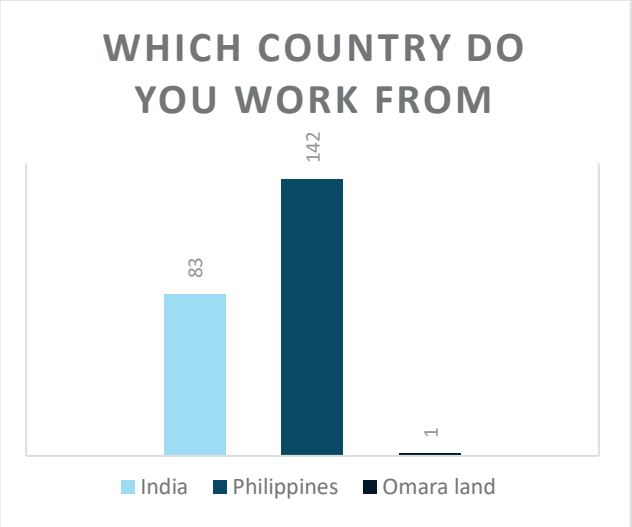
3.2.2 The annotators situation

Below the data collected regarding the annotators and their situation will be presented. The annotators are regarded as the main user of the solution and therefore of highest importance.

3.2.2.1 Location

This question was the one open question which it was possible to derive meaningful quantitative data from. The reason for not providing options in this question was due to secrecy, security, and integrity reasons, the annotators are shielded from knowing which other companies are partners to Omara. The results showed that the annotators were located in the Philippines and India, with most respondents in the Philippines, see Table 2. One respondent answered the county Omara is based in, but this is likely due to a misinterpretation of the question.

Table 2 - Annotators country



3.2.2.2 Language

Given that the GL is written in English it was deemed interesting to investigate how the annotators rated their skills in the English language. During the interviews, it became apparent that translation tools sometimes were used by annotators to translate parts of the GL. In Table 3 it can be seen that most annotators rate their English skills as average or above average. Whilst this points towards that the annotators feeling comfortable with the English language it can be hard to draw any conclusions regarding how good they are. Partly because they were asked to rate their own skills and partly because it depends on what they compared their skills to, e.g., did they compare their English skill to other people amongst them in their community or did they compare to a native English speaker? Furthermore, it can be hard to estimate what average means in this context.

Table 3 - The annotators rating of their English skills.

How would you rate your english skills?

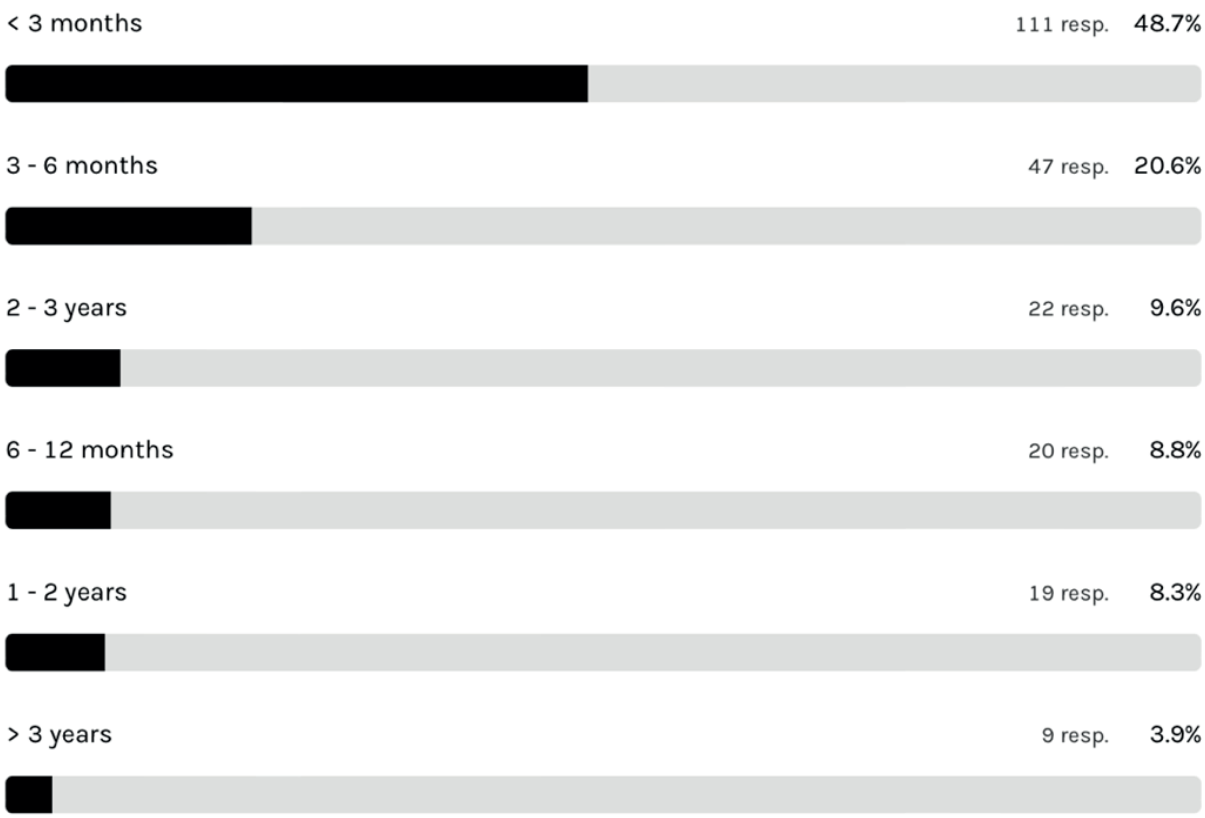
228 out of 228 answered

	Beginner	Below average	Average	Above average	Advanced	Mastery
Reading	0%	2.2%	32%	33.8%	24.1%	7.9%
Listening	0.4%	2.2%	33.3%	39%	18.4%	6.6%
Speaking	0.9%	7%	60.1%	22.4%	8.3%	1.3%

3.2.2.3 Experience as annotators

Approximately 70% of the annotators had worked less than 6 months for Omara, which indicates that most are relatively new to their platform and way of working, see Table 4.

Table 4 - The experience the annotators have working for Omara.



Furthermore, it was also asked whether they had worked for another company than Omara, where 100 answered no. Out of the people who only had worked for Omara 44% had worked for less than 3 months indicating that the entire task of annotating, in general, is very new to some of the annotators, see Table 5.

Table 5 - Table showing how long annotators that only have worked for Omara, has worked as an annotator for Omara.

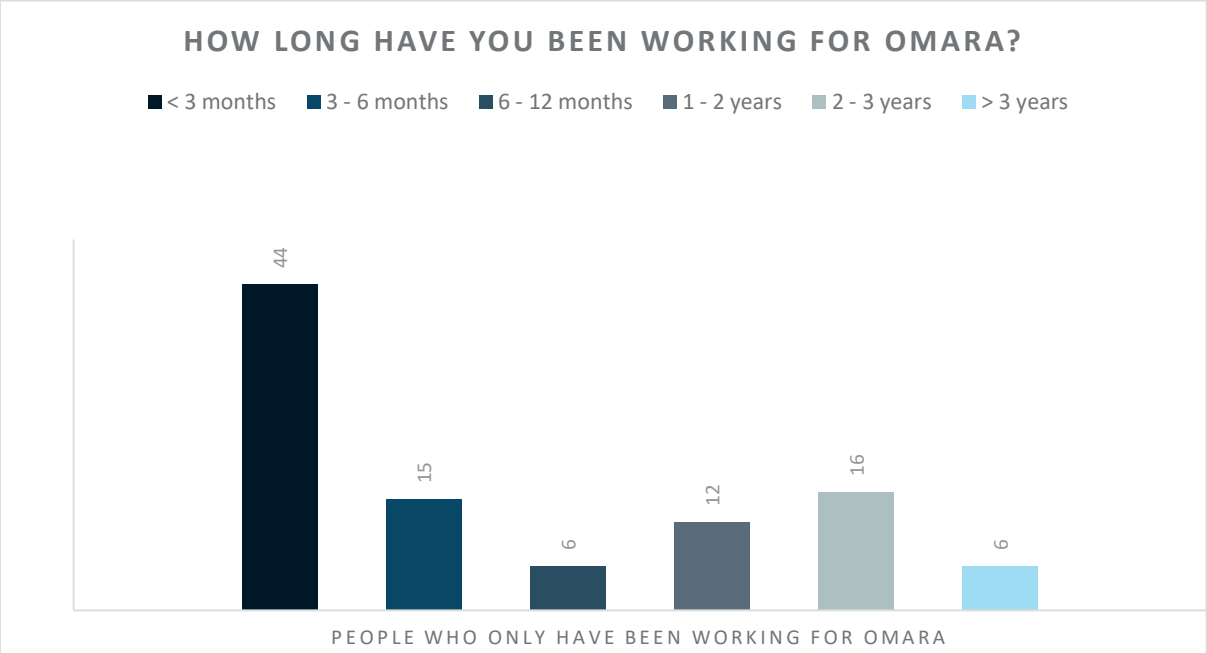
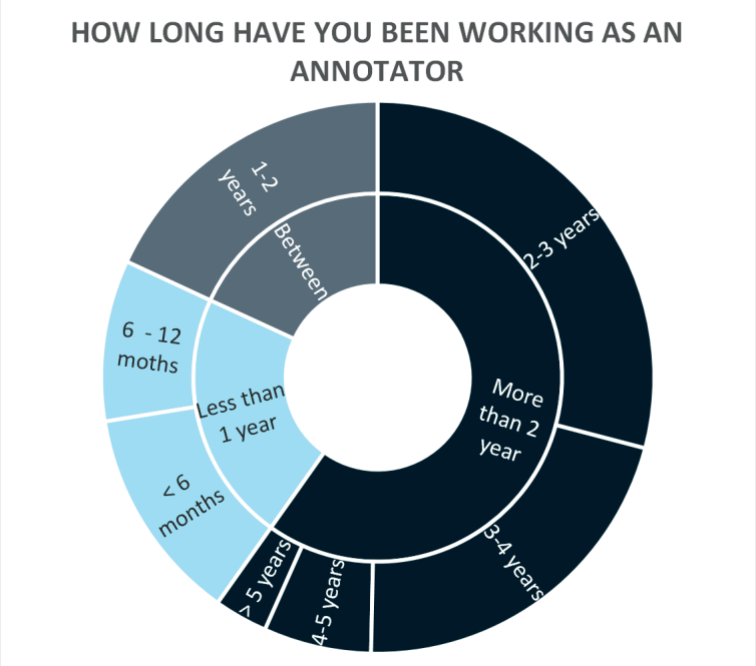


Table 6 - The experience the annotators have working for Omara or other companies

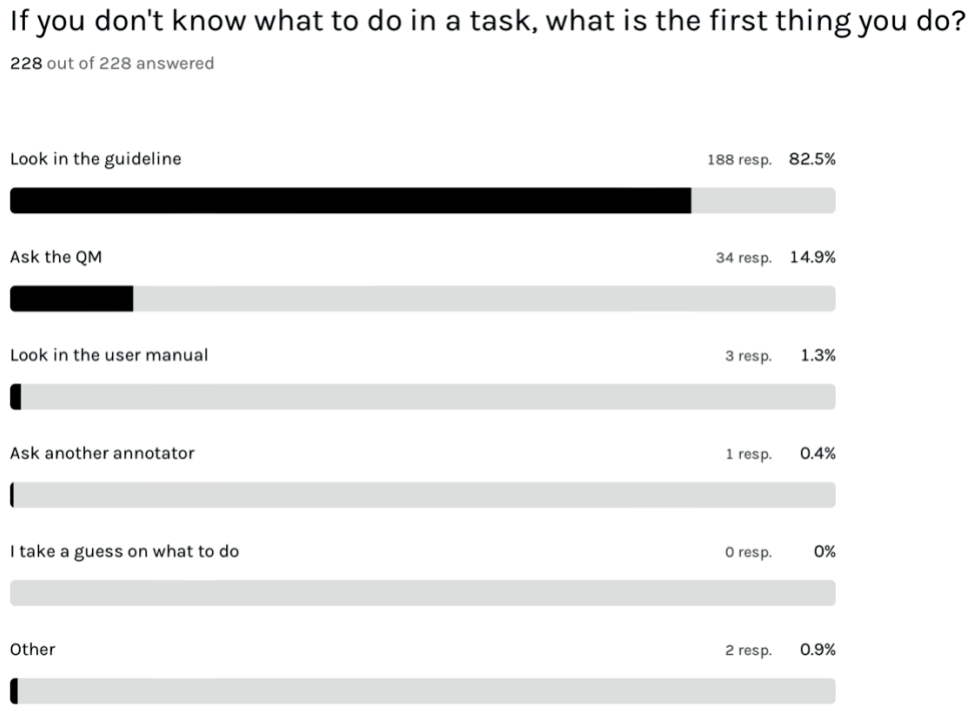


As can be seen in Table 6 a majority of the Annotators, who had worked for other companies than Omara, have 2+ years of working experience indicating that there is a widespread of competence across the survey participants. However, strengthening the fact that many of the annotators are new to the task and annotating it can be seen that 22% of those who have worked for another company than Omara still only has worked less than a year in total as an annotator, see Table 6.

3.2.2.4 When problems arise

It was deemed interesting to investigate how frequently the GL was used due to it being the main source of information. As can be seen in Table 7 a clear majority primarily turns to the GL when they are unsure what they should do in an annotation task. However, not all annotators use the GL as the first thing they do when they encounter a problem. Around 15% of the respondents said that they go directly to the QM (BPOQM) instead. This might point toward a hesitation to use the GL by some of the annotators.

Table 7 - Response of what the annotators do when they are unsure of what they should do



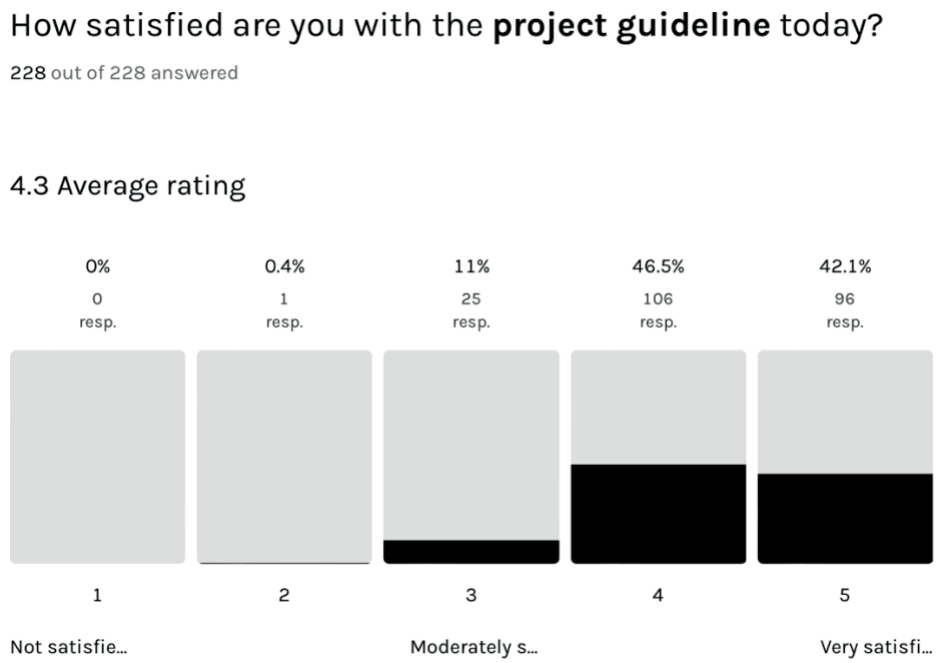
3.2.3 The annotators view on the guideline

As mentioned previously the guideline is the primary and, in many cases, the only source of information. Therefore, it is of high importance to receive information regarding the annotator's satisfaction with it and what rooms they find for improvements.

3.2.3.1 Level of satisfaction

As can be seen in Table 8, the annotators are in general quite content with the GL today. It shows that there is no widespread discontent amongst the annotators with the current solution. This is an important aspect to have in consideration when designing a future solution. However, it doesn't mean that there is no room for improvement.

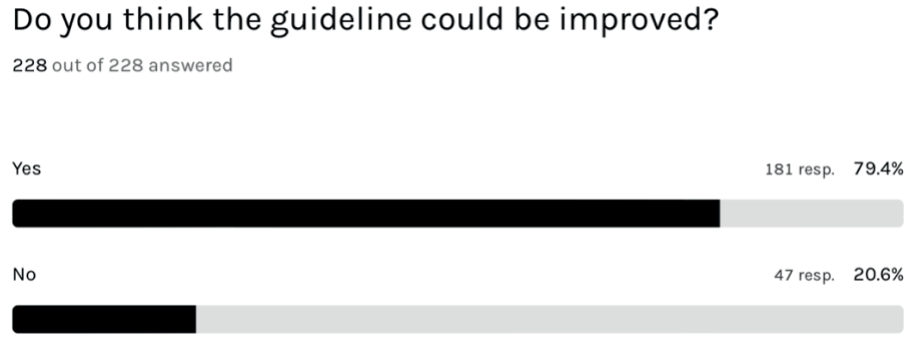
Table 8 - Table showing how satisfied the annotators are with the guideline.



3.2.3.2 What could be improved about the project guideline?

Even though most annotators were quite pleased with the GL, a clear majority still think they could be improved, as can be seen in Table 9.

Table 9 - Response when asked if the guideline could be improved



When asked what could be improved with the GLs a vast variety of answers were given. Some of the answers could be discarded because the answers were unclear or impossible to understand. For example, “the annotation” or “it helps me a lot”. The kept answers were compiled in a table where similar answers were grouped and counted. The most common suggestions (mentioned more than 4 times) were:

- More examples

- More information/details
- *Shortcut keys*
- Version history
- Sample videos
- Index
- More updates of GL (after feedback/error)

The suggestion which was the most reoccurring one was “more examples”. Almost half of the respondents suggested that the GLs should have more examples. Another common suggestion was that the GL should contain more information and more specific information to certain situations and details. More shortcut keys were suggested multiple times; however, this is more related to Omara’s app rather than the GL itself, hence it being in italic. Also, worth mentioning regarding updates of the GL is that some annotators implied that they wanted as few updates of the GL as possible. The reason for this was that the annotators felt as though with an increasing number of changes, an increased risk of confusion occurs. On the other hand, some annotators stated that they wanted more updates so that the GL contained relevant information, examples, etc. Both these aspects are therefore important to keep in mind simultaneously, that whilst some wishes that the GL is updated more frequently, others find it frustrating.

3.2.3.3 Multiple sources of information

From the interviews it became apparent that the IQMs and BPOQM sometimes creates addition documents that complement the instructions in the GL. This could be in the shape of summaries, newsletter, and communication via internal communication channels. The question was therefore posed in the survey whether the annotators receive other documents beside the GLs or not. The response from this question was that a majority of the annotator stated that they receive additional documents. However, in the following open-ended question, where the annotators were asked to provide examples of which type of documents there were, it became apparent that some of the annotators might have misunderstood the questions. Examples of additional instruction were given, such as: Summaries, best practice, video tutorial, key points. However, a lot of the answers given were in line with: Guideline updates, PDF, N/A, Google docs. This created an uncertainty whether the question had been understood correctly or not. The exact result from this question is therefore not presented here, as it could be misleading. What can be stated however, is that the annotators sometimes receive additional instructions beside the GL, but exactly how common this is hard to derive from the results from the survey.

04.

DEFINE

The goal of this phase is to converge the results from the discover phase into a concise description of the problem and to set up requirements for a future solution, which can be used in the upcoming phases.

Find insight

Build themes & clusters

State requirements

User de-
scription

4.1 Methods

In the following sections the methods used to define the problem will be explained. The goal of using these methods is to use the insights from the previous phase and compress them into a coherent description of the problem and ensure that the users' requirements are seen to.

4.1.1 Flowchart

In order to gain a deeper understanding for how the instructions are communicated from the customers all the way to the annotator, and how the information is treated and modified along the way, a flow chart was created. The flow chart was created from the information which was gained during the interviews with the IEs. The IEs also provided feedback and input during the creation process of the flow chart. This feedback was then used to iterate and refine the flow chart.

4.1.2 Affinity diagram

An affinity diagram is a way of structuring quantitative data from, for example, interviews, observations, and insights, by grouping extracts from the data and grouping them to find patterns and themes (Dam & Teo, n.d.). In this project, an affinity diagram was used to analyse the data from the interviews with IE, IQM, BPOQM, and the recorded interviews with the customers. The transcribed interview was printed, and quotes were cut out. Each quote was then analysed and placed with quotes that touched upon similar topics or expressed similar opinions. When needed, the groups were split into smaller groups or combined with other groups.

4.1.3 List of requirements & functions

A requirement specification is a compilation of requirements a product needs to meet, but it can also contain wishes that would be preferable if the solution met. These requirements can come from the user, customers, or other stakeholders (Ulrich et al., 2020). Its purpose is to guide the development phase and act as a reassurance that the concept created fulfils its purpose (Wikberg Nilsson et al., 2015). A list of functions is used to break down the functions needed into necessary or desirable functions within the solution. The functions a product needs to possess are broken down into three categories: main-, sub-, and support-function. The main function is mandatory for the solution to fulfil its purpose, in other words, the main function is the purpose of the solution. A subfunction enables the main function and the support-functions are important but not necessarily mandatory for the main function to fulfil its purpose (Bligård, 2015).

In this project, a list combining both requirements and functions was produced. Furthermore, the requirements/functions were weighted by importance accordingly:

1 = Good to have
2 = Somewhat important
3 = Important
4 = Very important
5 = Extremely important

4.1.4 Personas

A persona is a description of a fictive character to give a portrait of the target group that is more colourful than raw data and statistics. The foundation for creating a persona should be based on user studies, e.g., interviews, observations, surveys, etc., which is then used to create a character that illustrates the users' motivation, driving force, goals, and frustrations (Wikberg Nilsson et al., 2015). By creating a persona, the user data might become more relatable and can be used to communicate findings as well as be used during evaluations. In this project, several personas were created to evaluate concepts and help think from the annotators or other main stakeholders' perspectives. Four personas were created, one to represent the IQMs, one to represent the BPOQMs, and two to represent the annotators. The IQM and BPOQM personas were based on the interviews conducted and the annotator personas were based on the findings of the survey. The personas were written in a way that presented some background information regarding each persona and then in an interview format.

4.2 Results

This section will present the current situation will be described and the identified problems, based on the results from the affinity diagram and flow chart. These problems have then been condensed into a requirement/function list and personas.

4.2.1 Current situation

Using the flow chart and affinity diagram some overarching themes could be identified. A description of these problem areas will now be presented.

4.2.1.1 Communication of instructions

Due to trade secrets, the full version of the flowchart can't be presented, but a simplified version of it is presented in Figure 2. What can be seen in the figure is that it starts with the customers creating the instructions, the GL, which the IE reviews and iterates in collaboration with the IQM. It is then sent out to the BPOQM and annotators who starts training on the GL. Once the training is done and the annotators pass the examination the production can start. Within this chain, it's the GL that is the foundation of the information communicated. This means that the annotators are far away from the customers that writes the instructions and turn to BPOQM or sometimes IQM if unclarities exists.

In the flow chart two loops can be identified, one before production and one during the production. During the first loop, the GL is reviewed and revised by the IE with the help of feedback from the IQM. It also goes through a quality test with the annotators before going into production. The second loop, the one during production, is related to when the annotators have problems/questions. If questions appear, the first thing the annotators usually do is consulting the GL before asking the BPOQM. If the question is not resolved it is then raised to the IQM, IE, and finally customer until the questions have been answered, see Figure 2. If the IE deems an update is necessary, the GL is revised.

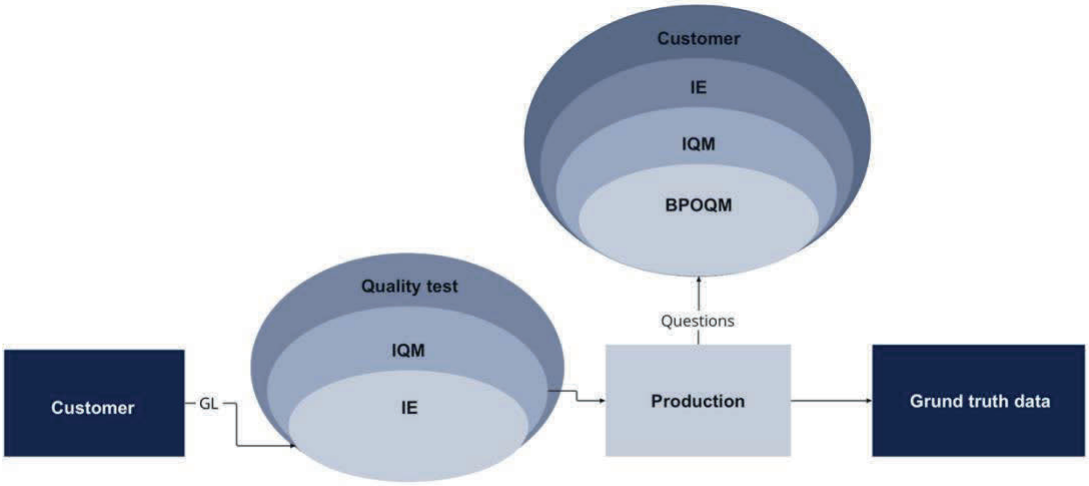


Figure 2 - GL flowchart from customer to delivery

Even though the GL is communicated to all the annotators, information can still come from different sources. Since the BPOQM is the annotator's closest manager most of the information they receive is from them and it is to the BPOQM that the annotators ask questions when they encounter problems. However, during the production process, the annotators might still get information from the IQM and IE. This isn't necessarily a problem per se, but different information sources could risk limiting the number of annotators that receive certain information, as well as leading to conflicting information, confusion and a scattering of the information. The way which this additional information is communicated differs, but some of the most common ways were identified to be:

- Summaries of important aspects of the GL usually written by the IQM or BPOQM.
- Newsletters where specific annotation situation and commonly asked questions can be addressed as well as information about any potential updates that have been made to the GL.
- Messages via platforms such as slack or discord, including both group chats as well as direct messages. Exactly which platforms that are used can vary, as different outsourced companies may use different communication ways.

4.2.1.2 Problems with updates

Updates are almost always required to be done in the GL, but the reason for the update can vary and so can the annotator's opinion about them as well. The updates can be divided into two main categories:

- Clarifying updates: These types of updates are made to make the information more understandable. This can be done by rephrasing an explanation, adding more examples, fixing inconsistencies, etc. What these changes have in common is that they are made to

improve the quality of the annotations by making it easier for the annotators to understand the instructions.

- Data changing updates: These types of updates change what is being annotated. Something that has been done in one way before should after the update be done in another way. For example, the values of a property could change, or the definition of a class. Or something that has not previously been annotated should, after the change, be annotated.

In general terms, the clarifying updates are seen as something good and the data changing updates as something bad, both by Omara and by the annotators. Since clarifying updates only should improve the quality of the annotation, it is something that everyone benefits from. The single most requested improvement on the GL in the survey was adding more examples.

" The guidelines should be complete in examples so that we will not ask our QM's anymore" (Annotator, Survey)

The reason for wanting more examples can be to clarify or exemplify a class/property, but it can also be to correspond to the actual data more accurately. Sometimes the examples can be from one country but the data that is being annotated is from another, meaning signs, road markings, traffic rules, etc., might differ. This might cause confusion since the examples don't fully match to the actual data. The same reasoning can be applied to data and examples from different seasons.

Since the PDF is a static document, it isn't possible to update only a part of the GL. Instead, a new version of the entire GL must be released, this however does not mean that the entire GL instructions will be changed, it can still be only a part that is changed but an entire new version of the GL must still be uploaded. This might increase the threshold of creating such an update. Another reason for these updates not to be made is that it takes time. When asked about how many of the GLs were rewritten one IE said:

"[...] what we should do is probably that we should rewrite more. But that we probably don't quite have the time to do so." (IE, interview. Authors translation)

The other type of updates, the data changing ones, are updates nobody wants to do but sometimes are necessary. The reason for these updates is that the data needs to be annotated in another way than what has been done up to that point. For example, the customer might have realised that the data properties need other values. From Omara's perspective this means that data that have been annotated once needs to be reannotated, so-called *corrections*. This might also cause some frustration among the annotators since they are required to relearn the changed parts of the GL and do things differently than they are used to. One annotator wrote in the survey when asked what could be improved with the GL:

" More clear rules about a certain project before actually doing the project.. because sometimee we get to start a certain project and the guidelines

suddenly change in the middle of doing the task. As an annotator it would be a lot helpful for us as a whole to have a FIX AND FINAL GUIDLINES before handing to the annototar. Because changing the guidelines while doing the task makes us confuse and our mind is sometimes remember the old gl that is prone in making a mistakes." (Annotator, survey)

4.2.1.3 Inconsistency

Since the GLs mainly are written by the customers themselves, the content and structure of the GL can vary a lot from project to project. One reason for this inconsistency is that the projects themselves vary a great deal in terms of what and how things are supposed to be annotated. For example, one project might be to annotate facial expressions, another to annotate vehicles, and a third to annotate road signs. While this makes it very hard to keep the instructions consistent, there are some common denominators. The GL presents a task, a certain number of classes with definitions regarding which object should correspond to what class, and properties with certain values. The fact that information regarding these aspects differs from project to project is hard to do anything about, but the way the information is presented could be questioned and possibly improved. Since almost all GLs are written by the customers it is very hard to keep things consistent, which points to a need for a new way of presenting the information.

4.2.1.4 Understandability

Since the GL is written by the customers, who might not always write them for the annotators, but rather as a definition of how they want the data to be annotated. It might lead to them being written in a language or way that the annotators fully understand, as one annotator stated:

"it most be improve the instruction given because sometimes it was confusing and also the QM also dont know how to answer that certain question of annotator" (Annotator, Survey)

While the IE tries to make the GL more understandable by changing and rewrite the instructions this isn't always successful, which becomes apparent in the quote above. One reason for this is that the time might not exist for the GL to be perfected. Another reason could be lack of understanding for the annotators situation and what they might find hard to interpret. What is obvious for someone who studied machine learning might not be as obvious for someone who isn't an expert in the area.

4.2.2 From problems to requirements

From the problem description it became apparent that there is room for improvement regarding how the instructions are presented to the annotators. The main way which the instructions are communicated is via the GL, but the fact that the annotators receive other documents besides the GL suggests that the GL is insufficient in communicating the instructions. Since the instructions are communicated in multiple ways the information might become scattered and hard to overview.

A new solution should therefore aim to make it possible to collect these different types of instructions into one coherent solution and/or make them obsolete.

Based on the findings from the user study, several specific areas of improvement could be defined. The single most requested improvement from the annotators was more pictures. Given that the instructions are communicated via a PDF, makes it hard to provide many images in the document whilst keeping the proximity which is asked for. It would be technically possible, but one potential risk is that the image would take up a lot of space and make the instructions inconvenient to read. Furthermore, more frequent updates, version history, and an index were suggestions which were mentioned several times.

4.2.3 Requirements

Based on the findings from the user study and affinity diagram, a list of requirements was created. The lists contain the requirements a future solution should fulfil to create a solution that helps to accelerate the annotators learning of the instructions. The requirements were then weighted to be able to use them in the theoretical evaluation of the concepts produced. The list was divided into sub-sections. These sub-sections were derived from the results of the affinity diagram. The sub-sections differ from the categories presented in the last section. The reason for this is to more effectively capture the different aspects a future solution must possess, and make sure that all the aspects and stakeholders have been included. In *main function* and *quality & speed* the overarching goal of the solution is defined. The sub-section *customers* exist in order to keep the effect of the solution's impact on the customers to a minimum. In the sub-section *content* and *technical aspect* most of the problems which was presented in the previous section is addressed. Finally, the sub-sections *communication*, *other*, and *wishes* addresses a couple of more general requirements which could not be included in the other categories.

4.2.3.1 Main function

The main functionality was divided into two sperate requirements:

- Solution shall communicate the annotation task
- Solution should contain instruction to fulfil the task

The first requirement states the solution communicate the actual annotation task, i.e., what should be done. The second requirement makes sure that the solution besides this also contains the relevant instructions for the annotators to successfully completing this task.

4.2.3.2 Quality & speed

An important aspect of a solution is that it helps the annotators to improve the quality of their annotations while remaining at an adequate speed. This can be seen as a main goal of the solution since better learning and understanding of the instructions can help improve both quality and speed. Two important requirements in this regard are:

- Solution should help minimise annotation time
- Solution should be effective and efficient to use during production

4.2.3.3 Customers

Since the main goal of this project is to accelerate the annotator's learning process, the focus of the requirement aims toward this. However, since it is the customers who write the GL a couple of requirements deal with the customers. The purpose of these requirements is to minimise any potential change in which the customers work today and interacts with Omara.

4.2.3.4 Content

The requirements in this sub-section make sure that relevant information is included in the future solution and is presented in a good manner. Some of the most important requirements related to this are:

- Should make the user aware of changes
- Text, including figure texts, must be readable
- Should express instructions rather than delivery

These requirements make sure that updates are easy for the annotators to notice so that no important information is lost. The way in which the information is presented in the solution must be easy to read and the information in the solution should focus on the actual instructions rather than what result the customer wants. Communicating what is desired does not necessarily mean that there is a lack of communicating how to achieve it, but the goal should be to communicate the instructions in such a way that ends up giving the desired result e.g., *how* to get the result rather than this is the desired result.

4.2.3.5 Technical aspect

The requirements that focus on the technical aspects of the solution aim towards making sure that the solution is easy to update, understand and find information in. It should be easy to update so that the annotators always can find updated, clarified, and correct instructions. Some of the most important requirements are:

- Should be easy to update/edit
- Should be digitally sharable
- Should be able to quickly find the desired content

4.2.3.6 Communication

The communication aspect of the requirement aims to make it easier for the annotators to ask questions and raise problems regarding the annotations and the instruction. It also aims to make the communication between the different roles more effective and efficient.

4.2.3.7 Other/wishes

Some of the requirements wasn't found to be possible to place in one of the previously mentions categories, but still deemed important to include in the list of requirements. A couple of requirements from these categories:

- Should have a high level of learnability
- Should minimise the amounts of documents

4.2.3.8 Complete list of requirements

In Table 10 a list of all the requirements is presented, including the ones that wasn't brought up previously in the chapter.

Table 10 - List of requirements with the requirements weighted.

Category	**Nr**	**Requirement**	**KJ refrences**	**Weight**
Main Function	1	Solution shall communicate the annotation task		Must
Sub Function	2	Solution should contain instruction to fullfill the task		5
Quality	3	Solution should help enhance quality whilst remaining speed	1.2, 1.3, 1.13, 1.14	4
	4	Solution should help customer to formulate what they want annotated/produced	1.4	2
	5	Solution should increase probability to cover most scenarios	1.2	3
	6	Solution should increase probability to cover most edge cases	8.6 13.3	3
	7	Solution should help minimise annotation time	6.2.3	4
	8	Solution should be effective and efficient to use during production	9.4, 1.14	4
Customer	9	Customer should not need to have insights in the annotation tool	2.1, 2.2	1
	10	Customer should not need to be in contact with many different people at Omara	1.5	2
Content	11	Should contain relevant examples for the data set being annotated	13.4 13.6	3
	12	Should contain many examples	12.4, 16.1	3
	13	Should make the user aware of changes	9.5	4
	14	Should only use simple English	7.10, 7.15, 13.7, 16.6	4
	15	Should express instructions rather than delivery	8.7	3
	16	Should have a consistent structure	12.1, 13.8	3

	17	Should have concise information regarding the annotation task and reduce amount of assumptions needed	12.2	3
	18	Lists and tables should be preferred over texts descriptions	16.9	2
	19	When referring to something it should be in direct proximity	16.11	3
	20	Text, including figure texts, must be readable	16.2	5
	21	Length of figure texts should be limited	16.3	3
Technicality	22	Should be easy to update/edit		4
	23	Should be digitally sharable	4.3 5.6 6.1.2	5
	24	Should be easy to comprehend regardless of technical background	Brief	4
	25	Should be able to quickly find the desired content	12.5, 16.4	3
	26	Should provide high resolution pictures	16.2	2
	27	Should communicate when a change has been made and the content of that change	13.9	3
Communication	28	Should allow annotators to feel safe to ask any question	7.9	2
	29	Should allow questions to be raised in an effective and efficient way to and between annotator, BPOQM, IQM, IE and customer	13.5	2
Other	30	Should have a high level of learnability	10.1 Brief	3
	31	Should use prioritisation of information and be concise	8.17 8.19	3
	32	Should minimise the amounts of documents	11.10,	3
Wishes	33	Should have the ability to include images, GIFs and videos	10.13 12.9 16.10	4
	34	Should encourage to use the solution rather than consulting BPOQM	9.6	3

4.2.4 Personas

The personas were created with the aim to give a good spread amongst them when it comes to personality traits and thereby cover many of the users. Since they were created in an interview format they could be presented with a more colloquial language, enhancing the appreciation of the characters.

The reason for there being one BPOQM and one IQM is due to them being deemed as highly affected by the new solution. They are the ones who need to learn it by heart so that they can answer difficult questions regarding it and be helpful and guide the annotators through the process.

Two personas were created to represent the annotator. They were created in extreme opposites, one very thorough that puts importance on the quality of the data and not as much on the annotating speed and one that prioritises speed and efficiency. They were used further in the idea generation process but were found most useful in the evaluation process. The personas do have somewhat conflicting requirements in the same way the customer has conflicting requirements, namely regarding speed and quality. In the way that the personas are created, they helped maintain a discussion regarding this so that neither speed nor quality must be compromised. Below the personas are briefly presented alongside some quotes extracted from each persona, these quotes are also fictive but there to enhance the characters. See A.1 Personas, for the full descriptions.

4.2.4.1 Annotator persona Bahvika

Name: Bhavika
Gender: Woman
Country: India
Language skills: English and native (Speak Tamil and English but for work almost only uses English)
Works for Gooseberry

Bhavika is an extremely thorough person. She is meticulous in everything she does and has therefore the highest quality rating in the entire company on annotations even though she only has worked for less than three months! There is no need to wonder if she will execute something with high results but her Achilles heel is time management. She asks a lot of questions so that she always feels 100% secure with what she's doing and would never just "take a guess" and often replies with "why" to explanations so that she can get the full picture. Knowledge is key for Bhavika and perfectionist is something she's been called her whole life, and she secretly likes it. She has only worked at gooseberry for a little less than three months but has already made a strong impression.

Bhavika regarding the GL:

> *"I use it for everything. I know it by heart. I read it so that i know exact quotes from it haha! No but honestly they are very importan[t…]"*

> *"I like them! I mean they can contain more examples and be more comprehensive but i like them. I like reading so for me these guidelines are good!"*

Bhavika regarding the goals and frustrations within her job:

> *"Being the best! No, but I guess it depends on who you're asking. My goal is to become IQM but I think my "role's" goal is to deliver high quality annotations within a certain timeframe"*

"That I feel as though I don't get enough information. I need all the information to do a good job. Also sometimes with updates, they can be so contradictory and confusing. Like, you just learnt it being one way and now you need to "relearn" it all and sometimes it is like "do it the total opposite".

4.2.4.2 Annotator persona Joshua

Name: Joshua
Gender: Male
Country: Philippines
Language skills: English and native (English is average, speaks Tagalog as native)
Works for Elderberries

Joshua is 25 years old and has worked as an annotator for about 3 years in total, and for the past 2 years, he has worked for Omara. When Joshua finds something that he likes to do or is passionate about, he commits fully to that purpose. However, when doing something which Joshua doesn't find as meaningful, his motivation sometimes declines. He isn't a lazy or sloppy person, but sometimes Joshua's quest for shortcuts can go a bit too far. This sometimes leads to a lack of quality in his work, as long as he doesn't find the work meaningful and rewarding.

Joshua regarding the GL:

"The guidelines are quite big and hard to find information in. I don't like to use and look stuff up in the guideline because it takes too long to find and that takes time away from my speed goal. Instead I usually trust my gut feeling instead of double checking the guideline, if I were to double check everything it would take a very long time."

Joshua regarding the goals and frustrations within his job:

"To annotate as many tasks that I can, while still staying above the quality goals. The speed is usually not a problem for me, but sometimes I can receive some quality feedback. "

"I would say that the most frustrating part is that it is so hard to find information. When I encounter a problem or an uncertainty it can be really hard to know how I can quickly find the answer. I like the summaries but they don't cover everything, so then I might have to open and look in the guideline anyway. This takes a lot of time and the correct information is hard to find. So sometimes I just trust my gut feeling instead and annotate the way I think it should be done. "

4.2.4.3 BPOQM persona Datu

Name: Datu
Gender: Male
Country: Philippines
Language skills: English and native (Fairly good at English, native Tagalog)
Works for Cloudberry

Datu is a driven, 29-year-old, ex-annotator that has shown off his skill which led to a promotion to QM 6 months ago. He is very proud of his work and wants to inspire his team to be the best team so that he can show off how good he is at his job as QM. He likes direct communications with his annotators, which has been hard for him during the times the Philippines has been in lockdown. He can come off as rather strict but is actually very soft and tentative when he is not under pressure. He has all the good qualities to be a good leader but is sometimes too goal-oriented and can then tend to forget the well-being of his annotators. When he starts a new project he is very soft and good at teaching his annotators the task and how to think. He uses very pedagogical approaches but when the deadline is closing in he can get a bit stressed and forget his softer side.

Datu regarding how and when he uses the GL:

> *For all purposes almost! I mean it is my job description and that's all I get so I have to basically use it for everything. But after a while you memorise things, this is something which is both good and bad since it makes it more efficient but it could risk generating more errors.*

Datu regarding the goals and frustrations within his job:

> *"Ah, well I take a lot of pride in what I do so for me, having a team that delivers good quality to Omara and within a certain time frame makes me feel fulfilled. I of course want everyone to feel good in their work environment, but the main goal for me and my team is to be the best team at Cloudberry, an opinion I think is shared among my team members."*

> *"Haha, when I feel like no one is listening. It's like you can give an annotator the same feedback on a problem but yet the exact same problem reoccurs over and over again. That is very frustrating. Also with the questions, I feel as though they are sometimes too lazy to check the guideline or something and ask obvious questions. That slowed down whilst we went remote then I felt as though I got too few questions."*

4.2.4.4 IQM persona Olena

Name: Olena
Gender: Female
Country: Ukraine

Language skills: English and native (Ukrainian native language and fairly good at English)
Works for Omara

Olena is 32 years old and has been working for Omara for almost as long as the company has existed. She started working as an annotator for Omara around 3 years ago and she has then advanced to the role as an IQM, a role she has had for about 1,5 years now. As a person, Olena is structured, ambitious, and to the point. She isn't afraid to tell people her opinion and she doesn't beat around the bush. Her ways of communicating might come off as harsh sometimes, but she doesn't mean anything by it, it's just the way she communicates. When someone has done a job that doesn't meet the expectations, she lets the person know, but she also makes sure to tell people when they have done a good job as well. If there's a problem, she makes sure to go to the bottom with it. There's is always a reason for why things aren't working, and she will make sure it's found.
Olena regarding how and when she uses the GL:

> *"I use it to make summaries and communicate the instructions to the BPOQMs and annotators. It is used as a basis for our work and when something is unclear it is always the guideline that is first consulted… "*

Olena regarding the goals and frustrations within his job:

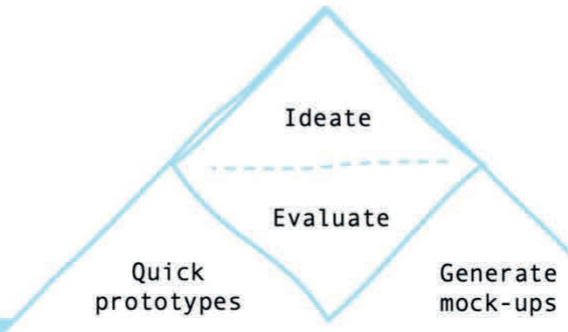
> *"The goal of my job is to make sure that the data that is being delivered is of sufficient quality by making sure that the annotators reach their quality and speed goals. I do this by helping them understand the guidelines, doing corrections and golden annotation, giving feedback to the annotators about speed/quality and pointing out errors which have been made…"*

> *"…the annotators keep making the same mistakes over and over again. We try to send them feedback on what they have done wrong, but the same mistakes keep occurring. Maybe they don't read the feedback, or maybe they have forgotten it. But it is annoying when you see the same mistakes repeated. Sometimes I don't think the annotators read the guideline that closely."*

05.

DEVELOP

In this phase all problems defined, and all requirements generated are aimed to be solved with one idea of a solution. The methods used to generate these ideas which then are combined into concepts will first be presented followed by the methods used to evaluate the created concepts. Furthermore, the results of these methods will be presented.



5.1 Ideation and evaluation methods

Introductory, in order to follow the process and understanding how concepts were created, the methods of the idea generation and evaluation phase will be presented. As a product development phase it was strived to generate many concepts to increase the chances of generating a concept which fulfils the requirements in a more efficient and effective way than others before introducing screening methods. It can be referred to as the product development funnel (Wheelwright & Clark, 1992) which, similar to the double diamond, is convergent in its evaluation phase that could be viewed in the same way as the screening, see Figure 3

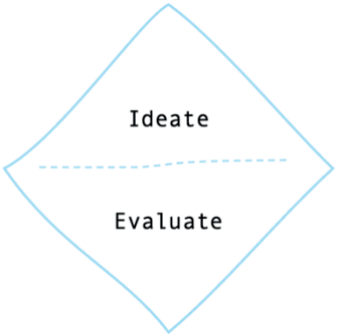


Figure 3 - Divergent Ideation and convergent evaluation

5.1.1 Brainstorming

The first ideation method used in the project was brainstorming. Brainstorming is a method which can be used to create a large quantity of ideas by generating ideas in a group (Osvalder et al., 2015; Wikberg Nilsson et al., 2015). The method is executed by allowing all the participants to come up with and present ideas, which are then written down (Wikberg Nilsson et al., 2015). During the brainstorming sessions criticism should be avoided, ideas can be combined into new ones and the goal is to create as many ideas as possible, preferably ideas that are "outside the box" (Osvalder et al., 2015; Wikberg Nilsson et al., 2015).

Before the brainstorming session started themes were created to help keeping focus and generate as many ideas as possible. The origin of the themes are the requirements, see Table 10, where the question "how?" was asked going through the requirements. They were then formulated and adjusted in a way that allowed for serval requirements to fit under each theme and to facilitate an ideation process. The themes generated were the following:

- Main communication method
- Communicating examples
- Make the user aware of changes
- Find the desired information
- Allow questions to be raised effective and efficient

Not all requirements fit under these themes but due to upcoming methods being dependent on them this was deemed as a good starting point, where the annotator is in focus. It was also realised that the themes could be viewed as sub-functions and will henceforth be called such.

A brainstorming session for sub-function was executed and the ideas were written down on a spreadsheet. Each brainstorming session regarding the themes had a duration of 7 minutes due to 5 minutes being estimated as too short to become creative enough to be allowed to think outside

the box, and 10 minutes deemed as to long for keeping up motivation throughout all sessions. At the end of all sessions between 9-12 ideas for each of the theme had been generated.

5.1.2 Morphological matrix

A morphological matrix is normally used by breaking down the main function into sub-functions. Then many possible solutions are defined to each sub-function. Once a matrix is created with the main functions listed in the first column and all solutions in matching rows listed in the other columns different possible concepts are explored by combining one solution from each sub-function until all possible combinations are explores (van Boeijen & Daalhuizen, 2010).

This means that the number of ideas created that then needs to be explored and evaluated quickly can become of great magnitude. Therefore, a simplified version of the method was used in this project. Instead of exploring each single possible solution, the different combinations were freely chosen, with the agreement that all solutions needed to be included in at least one concept and that a balance between concepts deemed feasible and concepts deemed more visionary was sought to be maintained. The sub-functions used in this method are the ones mentioned above in 5.1.1 Brainstorming and results from the brainstorming sessions were used as the defined solutions.

5.1.3 Solution Sketching

Solution sketch is a method which is developed to be used during a design sprint, developed by Google. The method is used to develop or combine ideas based on previous ideation results (Google, n.d.). Each team member develops two ideas on their own, based on previous ides, which is more fleshed out and of higher detail (Google, n.d.). This produced an additional 4 ideas.

The reason for using this method was to make sure that the solution space had been thoroughly explored and to be able to take a step back and ideate more freely than what the morphological matrix allows. The ideas were created by using simple sketches and/or text descriptions, in the same manner as the description of the previous ideas. The level of detailed of the ideas were deliberately not too high, in order for them to have the same fidelity as the previously created ideas and thereby possible to compare in the evaluations.

5.1.4 Six thinking hats

Six thinking hats is a method developed to aid thinking outside one's own perspective and instead think from specific mindsets and attributes correlated to a hat (de Bono, 2000). The method can help to systematically go through an idea, reduce conflict among the participants and see the ideas from a new perspective (Wikberg Nilsson et al., 2015). The method can be used in several different ways, but one way which is suggested by de Bono (2000), is the sequence use. This means that all or some of the hats are used and all the participants uses the same hat until the session is done. According to de Bono (2000) and Wikberg Nilsson et al (2015) the hats colours and their correlating attributes are:

- Red hat: Deals with emotions and feeling. How do we feel?
- Black hat: Cautious and identifies problems. What could go wrong?
- Yellow hat: Positive and sees the benefits with each idea. What is good?
- Green hat: Creative and sees new solutions. How can we do this different?
- White Hat: Focuses on facts and is objective. What do we know?
- Blue Hat: Is in charge of the process, organisation, and control. What is the purpose/goal?

The hats used in this project are the top four. Since there is a lack of facts regarding all generated concepts and due to it being bigoted to bring up only the facts regarding some concepts it was deemed as better to discard the white hat. Another discarded hat was the blue hat. This hat was somewhat used prior to the session to set up rules and goals of the session but not in the concept evaluation process.

During the evaluation each and one of the 13 concepts were evaluated with 3 minutes of evaluation time per hat. The thoughts and opinions regarding each concept were noted down and after the session was over the notes were analysed. From the conclusion 6 concepts could be eliminated. The reason for these decisions varied, but in general the concepts were eliminated because they would be too hard to implement or was deemed as to not have great enough potential to solve the problems in an effective way.

5.1.5 Pugh-matrix

The 7 remaining concepts were evaluated using a Pugh-matrix. A Pugh-matrix is a concept selection method which can be used to select concepts based on requirements. To use the method a matrix with the requirements on one axis and the concepts on the other needs to be set up, the concepts is then rated in correlation to if they solve the criteria better, worse or roughly the same as the concept which it is compared to (Muller, 2009).

During the first iteration of the Pugh-matrix the concepts were compared to the existing solution (the GL). As mentioned in 4.1.3 List of requirements & functions, the requirements were rated between 1-5. This weighted number was then multiplied with 1, 0 or -1 depending on how the concepts stood to the current solution. After all the requirements had been rated one concept had the highest score. This was then used in a second iteration where the winning concept was used as the solution which the other solutions were compared to in order to find out whether it actually is deemed as the best concept, see A.3 Pugh-matrix.

During this session some concepts were deemed similar enough to combine. By keeping the attributes within the concepts that gave the higher results in the matrix 4 concepts were combined into 2 concepts instead. Furthermore, two concepts were eliminated, one due to it performing inadequate in the matrix and one since it entailed a different media of operation which was deemed as impossible to require from the annotating companies. In conclusion this method evaluated, improved, and eliminated concepts which resulted in moving forward with only three concepts.

5.1.6 Mock-ups

Mock-ups are static yet a realistic way of showcasing what a product would look like, it is a quick and easy way of prototyping, whilst a real prototype would be of higher fidelity and interactive (UXPin, n.d.). Mock-ups or quick prototypes are of great value when something needs to be visualised in order to test the concept, to see whether it is feasible and to communicate its function (Ulrich et al. 2020).

In this stage of the project mock-ups were needed in order to portray and communicate the concepts in the concept evaluation. By using Figma some wireframes were created. This mock-up stage also became like a small idea generation since smaller undefined icons, positions and similar needed to be clarified and defined before presentation. Several suggestions on how to portrait, for example the icons, were created and then decisions were made based upon the knowledge gained in chapter, 2 Theoretical framework.

5.1.7 Evaluation workshop

According to Ulrich et al. (2020) one can use evaluation workshops to evaluate opportunities. They continue to describe these workshops as short, sequenced presentations where the participants vote on ideas they perceive having the greatest opportunity. They also suggest that the voting should be conducted simultaneously by all participants and in silence to avoid influencing each other.

In this project the evaluation workshop based upon what Ulrich et al. (2020) presents whilst combined with a partial presentation for the company. The participants invited to the evaluation workshop was two UX/UI designer, who also was the company supervisors for this project, an IE, and one participant who works with the quality of the data. Since the participants all had great knowledge regarding the subject and how feasible the concepts would be, the idea was to have an open discussion and evaluation of each concept to share knowledge and thoughts between participants. After evaluations of each concept a final discussion and general evaluation was held where participants were invited to share other ideas to help evolve the concepts.

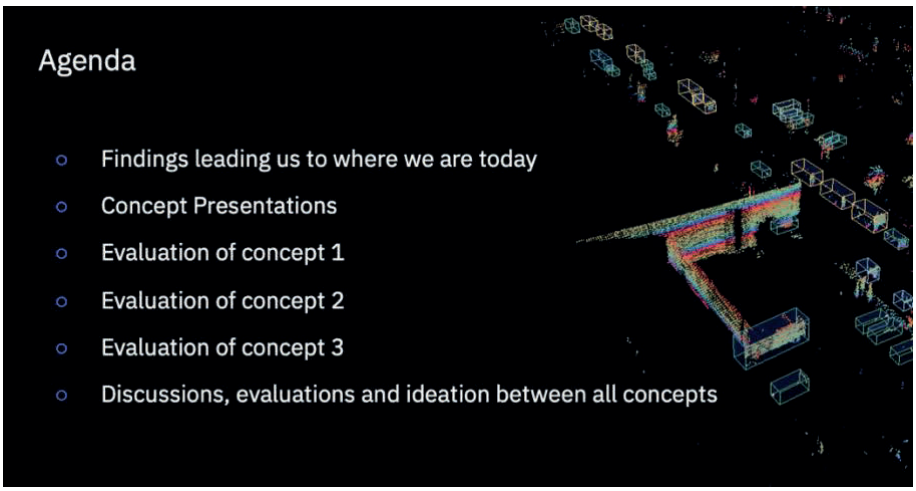


Figure 4 - Evaluation workshop agenda.

The evaluation workshop therefore had the agenda to be seen in Figure 4. Prior to the workshop the participants were handed the personas and list of requirements so that they all could have share point of views and be aided in their discussion.

5.1.8 SWOT

A SWOT is an analysing tool created by Albert Humphrey with the purpose to reinforce a business strategy by analyse and asses the organisation regarding four areas. Those areas are strength, weakness, opportunities, and threats. The strength and weakness area of the tool refers to internal factors i.e., what happens within the organisation whilst the opportunities and threats refers to external factors that could impact the organisation i.e., what outside impacts the organisation such as competition, trends, laws etc. By combining both internal and external factors it is easier to find out if the benefits risk being overshadowed by the disadvantages and how to take positive actions (British Library, n.d.).

A SWOT-analysis can however be applied to many other things than just an organisation. In this project it was used to lists internal and external pros and cons regarding the three concepts. Such as how realistic it is to implement, how innovative it is, who will realise the concept, etc. It used arguments from the Evaluation workshop to act as a foundation to the ultimate screening where only a final concept makes it through.

5.2 Results

In this section the results from the idea and evaluation methods presented above will be explained. Some of the early concept will only be briefly presented whilst some of the concepts remaining after evaluations will be presented more in depth.

5.2.1 Early concepts

The brainstorming and morphological matrix lead to a generation of 12 concept, see A.2 Morphological matrix. After the concepts had been generated, they were quickly evaluated by discussing the concepts one by one. Since some of them used the same main communication method they were deemed to be too similar for them to be useful in the upcoming evaluations. Hence, some of them were combined into one concept resulting in 9 concepts being left after the morphological matrix evaluation, see Figure 5.

Main Communication method	Examples communicted	Make user aware of changes	Find desired information	Should allow questions to be raised in an effective and efficient way
Everything embedded in the tool. Adapts info to where you are in the platform	know what kind of example should be communicated	red text, "lingon berries"	Filtering fonction automatically, searchbar to find outside what is defined by the platform	forum in platform, cutting tool to take screenshots and sent to desired person. Anonymous and with @ person
Webpage, Interactive digital document, "roadsign style"	Pictures	Notifications in platform, marked text, see changes	Filtering function	support errands
a person sits and looks through all the pictures and categorizes them then there are categories all annotators annotates in	tutorials	onboarding	reward system "ask a friend" friend who answers will be rewarded	chatt
AR	popup with "examples"	time frame you have to look at text	Presents only the information you need	chattbot
PGE - Personal GL expert	separate screen with exampels	quiz	simpler communication method	hotline
AI "chatt bot"	searchable image bank	read receipt	"maybe this was really what you were looking for"	live call with screens
PDF	Pictures	versionhistory	search function, clearer paragraph breaks	forums - open anonymous questions, others annotate can answer as well
Tables	"lexicon"	"disabled annotation platform"	index	faq
Voice	separate screen with exampels	irritating sound	"hey siri"	support errands

Figure 5 - 9 Morphological concepts

Since the Solution Sketching resulted in 4 new concepts a total of 13 concepts had been generated before going into the evaluation made with Six thinking hats and the Pugh-matrix. To concretise the concepts, explanatory texts were written for each concept and where it was deemed necessary explaining sketches was made as complement.

5.2.2 Producing 3 concepts

After refinements and evaluations using the methods Six thinking hats and Pugh-matrix the screening process had taken what once was 13 concepts down to only 3. These are in more detail and of higher fidelity then what the concept ideas once had. This section will present the three concepts left in the screening process together with the outcome of the SWOT analysis regarding these.

5.2.2.1 The PDF

This concept is similar to the already existing concept, since it's PDF based. The tweak here is that a general GL structure is generated so that all customers GLs shares structure, which could cause less confusion for the annotators. An example of this cannot be provided as it would look too similar to the existing solution, and thereby revel trade secrets.

By looking into many existing GLs, a structure that is deemed as user friendly and logical will be generated. The annotator should instinctively know where to find certain information and therefore always know where to look and furthermore all projects should look the same. Although, important to keep in mind is that not all annotators might share the same instincts so it might still differ. However, with time and learning the "new structure" the annotators should be able to locate all information quickly.

The structure will be transformed into a template that can be sent out to the customers to follow if they want to do it themselves otherwise, they can pay Omara to convert their GL into the Omara template for them. By having a template, the likelihood of missing information will be reduced, and the GLs will be written in a more instructing way rather than expected delivery way.

This concept is easy and cheap to implement and wouldn't be a big transformation for the annotator and their workflow. Other areas that generate doubts regarding this concept are:

- Not innovative
- Might not make any difference
- What is a good structure?
- Would not affect the speed to much since annotators still must go through the entire GL

After the Evaluation workshop a SWOT analysis was made, and the results can be seen in Figure 6.

Figure 6 - The PDF SWOT

5.2.2.2 The Eagle Eye

The idea behind this concept is inspired by user documentation sites and the navigation used with dropdown menus. The thought is to generate a good overview of the entire instructions so that it would be easier to navigate within the instructions. It's based upon prioritisation of information so that the main headings initially will be shown, and the annotator would know which heading to explore. Once a heading is chosen, the annotator enters the next stage where subheadings will be shown. The annotator can also choose to open all heading and then it is more similar to the existing GL. It would be website based but accessed through the platform where the GL is found today.

Once a change is made to the instructions an alert will be made in the form of a pop-up. A notification badge will also appear where the link to the instructions is situated, and this badge will disappear once the change is read, see Figure 7.

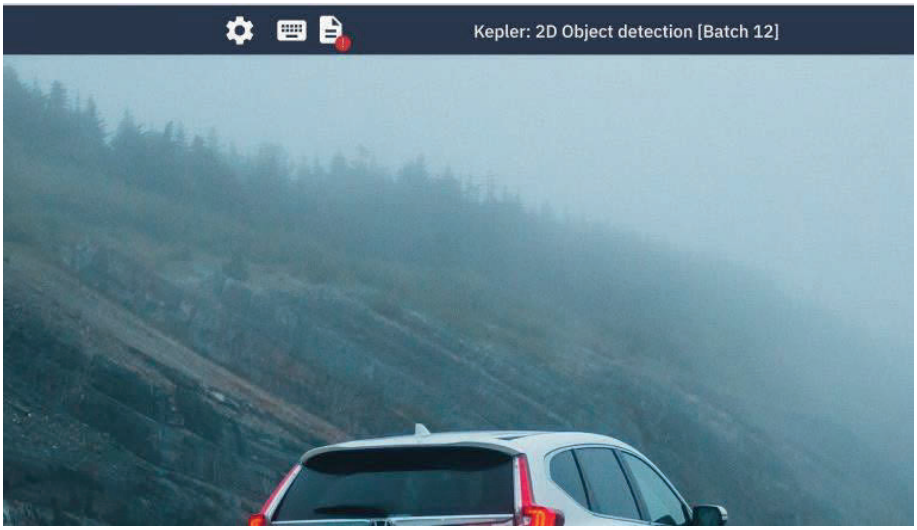


Figure 7 - Notification badge in platform

When the users open the instructions, badges will show underneath which headings the changes occurs but there will be no direct link to the change, see Figure 8 & Figure 9. This is to make the annotator highly aware of exactly in what area the change is. All these notifications will increase the learning and navigation once changes have been made. In addition, to decrease frustration, a version history will be available to choose to be shown.

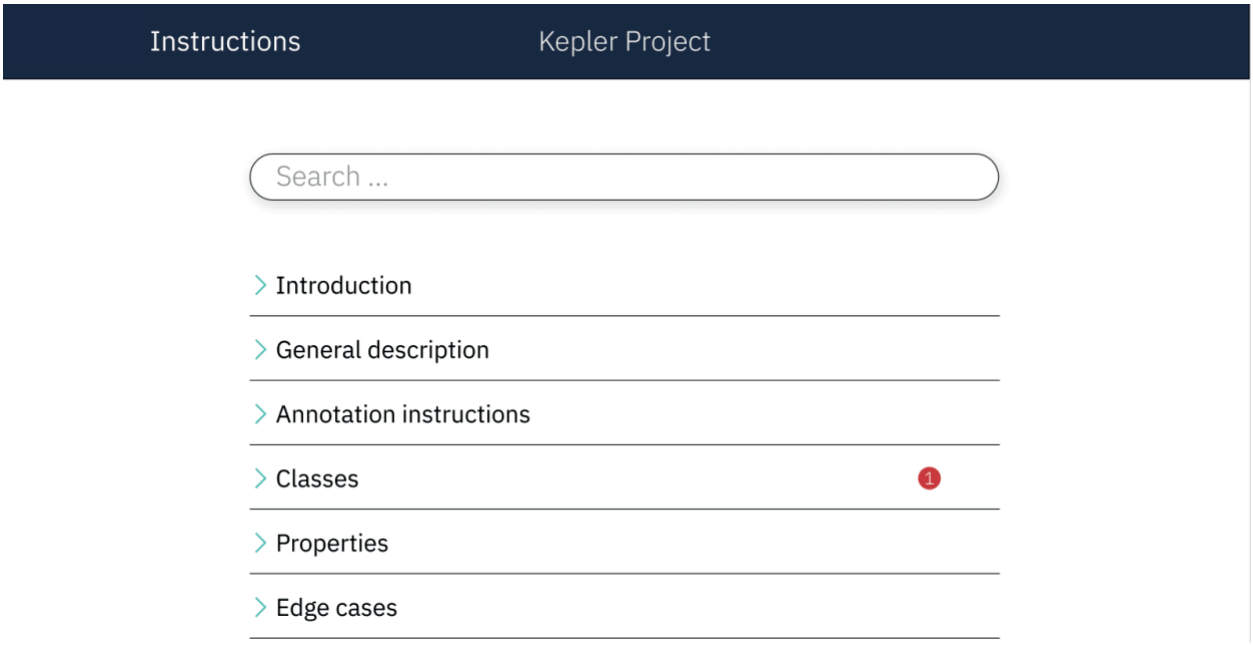


Figure 8 - Home screen of Eagle Eye with notification badge

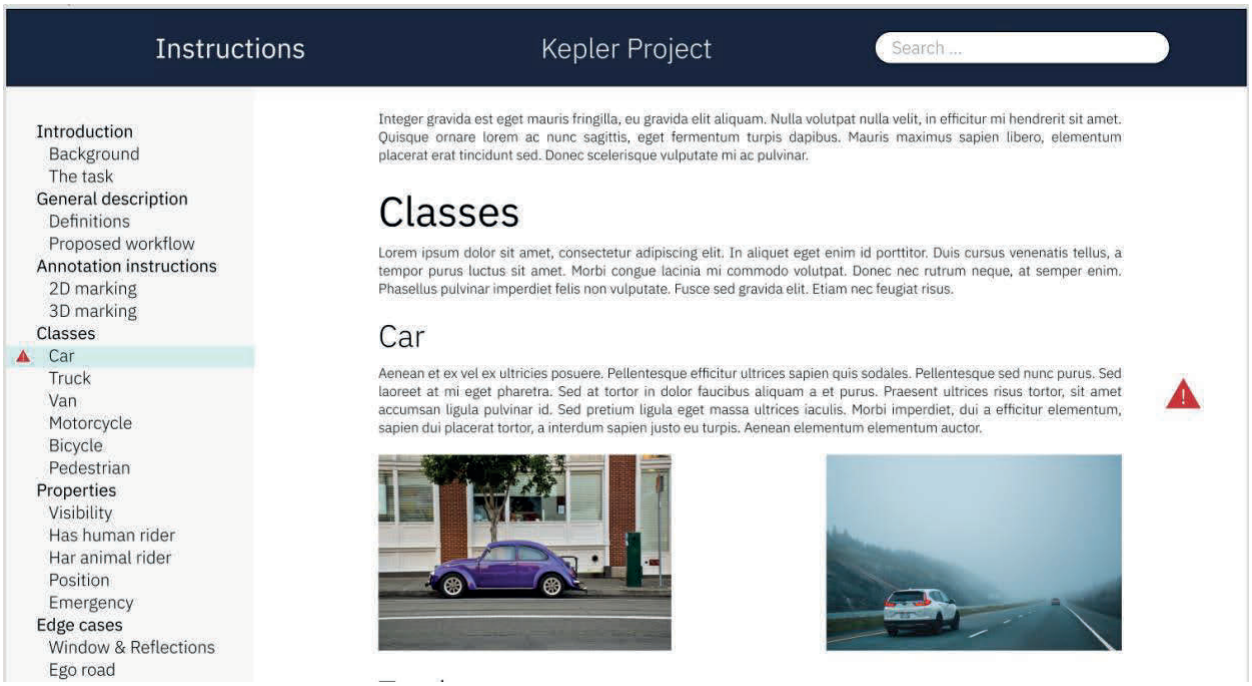


Figure 9 - Notification badges in Eagle Eye

The instructions would be easy to update and minimise the risk of the annotators missing the updates. For Omara this concept would be rather easy to implement and not give too much extra

work. Someone will have to turn the GL handed in by the customers into this interactive document and then further update it when changes occur.

This concept wouldn't impact the customer side much since an Omara employee would be the one transforming the GL into this interactive online document. It provides better overview of the instructions and reduces risk of error with changes, it is interactive, and the user can decide how much information should be displayed. Some risks with this concept would be that someone would have to work fulltime with this, moving GLs into this interactive platform, and working with updating them. This concept also demands more interactions and many keystrokes. The results of the SWOT analysis where Omara employees had inputs can be seen in Figure 10.

Figure 10 - Eagle Eye SWOT

5.2.2.3 The Tour guide

The third concept is maybe the most technically advanced concept. It is highly complex and relies upon the Omara platform. The concept is based on prioritisation of information as well as platform interactions. Like the first concept, the PDF, this concept can't be shown as it would reveal trade secrets.

Depending on which step the annotator is working on in the platform, information is displayed. For example, suppose no class is chosen, then the information level the annotator would receive is regarding class. When class is chosen, then information regarding properties etc. will be displayed. This means that the information displayed always helps the user to conduct the next step in the process.

To be on the safe side and for the training process, the original GL will still be available to study and the information integrated is there more as support during the annotation process. Hence, this concept is focused on increasing the speed and quality of the annotator's work rather than the learning process.

The idea behind this concept is to only give the annotator the information they need in an effective and efficient way, so as little time as possible should be spent on looking up information. From Omara's side this concept requires rather lot of work, since the GL must be integrated into the platform and the right level of information needs to be determined for each step. An idea is that the customers themselves could go into a customer viewpoint and add the information themself. This would however increase the risk of the customer misunderstanding and display information in insufficient or incorrect way. The best situation would be to have an Omara employee set up the platform and ensure that the right information is displayed at the right place, but it would also affect the customer prices. If it can be shown that the quality of the data is increased that the customers would be likely to pay extra, Omara ensures.

Similar to the previous presented concept, when changes to the instructions are made a pop-up is generated. The new GL is uploaded in the same way as today, but the real new function of this concept is that a badge will be shown in the sidebar in the annotation platform regarding the area of change. That badge will be there until the change is read to ensure that the user has read the change. Some benefits with this concept are that it should improve speed and quality whilst being intuitive and innovative. It should also reduce the risks of errors or missed updates and deemed as a safe bet due to the old solution still existing. Some disadvantages with this concept are that it is time and resource consuming and hard to implement. Also, when there are changes these needs to be updated in two separate places, both GL and in the platform bar. Another problem which came up during the Evaluation workshop was that the notification badges and the already existing alert for when a property isn't chosen would be interfering since the icons used would be extremely similar. Also, the tool tips including the annotation instructions could be interfering, the user might think other information would appear then what actually is displayed, which would be frustrating

for the annotators. Analysis of this concept together with what occurred in the evaluation workshop is presented in Figure 11

Figure 11 – The Tour guide SWOT

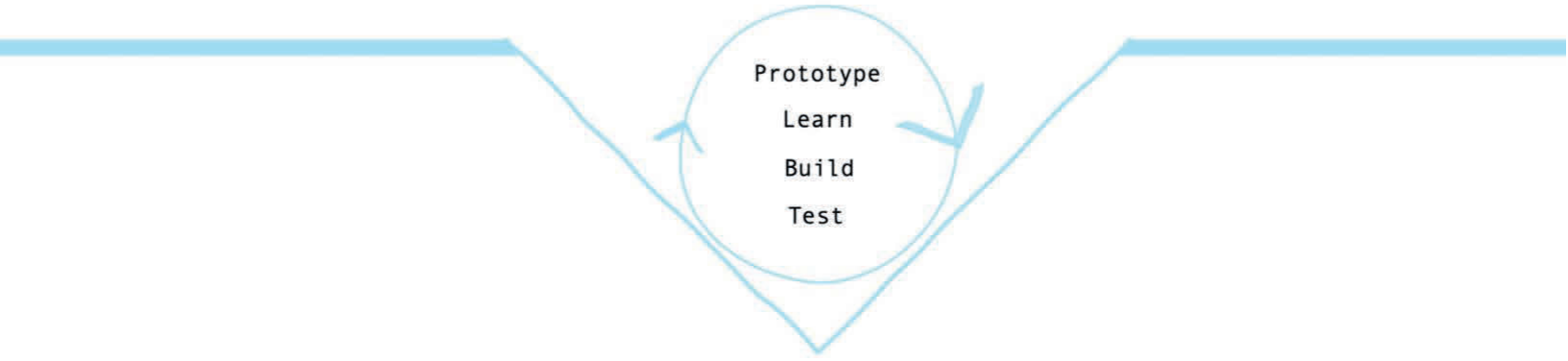
5.2.3 Selecting concept

After the rigorous idea generation and evaluation, the final screening had to be done. Both advantages and disadvantages with all concepts can be read under respective section in 5.2.2 Producing 3 concepts. All concepts were deemed as realistic to implement and could be introduced as suggestions of improvement of the current GL. However, the uncertainties of “The PDF” made it the least favoured during the Evaluation workshop and the final SWOT further indicated for it to be screened out of the process. How implementable the remaining two concepts are further acted as a foundation to the final screening. “The tour guide” is a concept where many potentials could be found, however there were many unanswered questions regarding how well it actually would work that interfered. It was also deemed to function better if another solution to the original GL was implemented prior. Hence, “the Eagle Eye” made it through the final screening process.

06.

DELIVER

This chapter will present the phase where prototyping and testing are performed. It is a converging phase where the final solution is refined and produced. The methods used to refine and validate the concept will be presented, followed by a section regarding the results from these methods.



6.1 Methods

Below will the methods of the phase “deliver” be presented. These are methods conducted in order to refine, test, and produce a final concept.

6.1.1 Creating structure

Since all projects have different instructions, it is hard to find a structure of the instructions which fits every single project. However, an attempt was made to create a basic structure that could be applied to as many projects as possible. To do this 14 GLs were picked at random and broken down, where the content of each GL was summarised on post-it notes. The post-it notes were then rearranged to find common denominators. The result was 5 overarching categories: Introduction, Annotation instruction, Classes, Properties, and General examples. The vocabulary used to describe these different categories differed in the different GLs. For example, instead of properties, the word attributes could be used, and classes were sometimes described as labels. Furthermore, every GL didn’t necessarily contain information about each one of these categories. But in these cases, the category could simply be disregarded.

6.1.2 Prototyping

The purpose of prototyping can vary depending on which design phase the prototype is created in. According to Benyon (2010) a prototype is a partial, but clearly defined representation of a design, with the main purpose being the ability to evaluate a concept. Lim et al. (2008) add another perspective to the subject and argues that prototyping, besides being part of evaluations, can be a method to explore the solution space. They argue that prototypes can be seen to discover new ideas and refine existing ones (Lim et al., 2008). The level of detail and resemblance to the final product is called the level of fidelity (Benyon, 2010). While lower-fidelity prototypes might be suitable earlier on in the design process, high fidelity prototypes are more useful when performing detail evaluations (Benyon, 2009).

The goal of the final prototype created in this project was a combination of the approaches presented above. The final concept needed to be further iterated to explore the solution space and the ideas presented in the evaluation workshop needed to be incorporated into the design. In addition, it needed to be refined enough so that it could be used in an evaluation. Before the actual prototyping started sketches were created to define what elements should be a part of the solution, their placement, and some basic interactions. When the basic layout had been agreed upon the prototyping started. Like the mock-ups, presented in 5.1.6 Mock-ups, the prototype was created in Figma. During the development of the prototype several shorter ideation sessions, mostly using brainstorming as a method, were held to solve problems and questions that arose. For example, how the interaction with a menu item should work or how the version history should be displayed. Some of these ideas were simply discussed verbally or explained using simple sketches, whilst others were prototyped in Figma.

6.1.3 Usability testing

Usability tests are according to Preece et al. (2015) tests that can be done to measure how usable an interface is when performing the actual tasks for which it was designed. Moran (2019) specifies this further by stating that the most common goals usually include identifying problems, potential opportunities that can be improved, and learning the behaviour and preferences of the target group. Usability tests can be performed in many ways, one approach, which is the one used in this project, is in a controlled setting or lab environment. In these types of tests, environmental factors are limited, and the facilitator can control what task is to be done as well as how and for how long the user should perform each task (Preece et al., 2015). Controlled tests are suitable for testing a specific hypothesis and collecting quantitative data such as time and errors (Preece et al., 2015).

The test was carried out with 4 participants. Since the information in the prototype due to secrecy reasons couldn't be shared with people outside the company the participants were employees at Omara. All the participants had limited knowledge of the annotation task. This selection was made to make sure that the participants used the information presented in the solution and not relied on previous knowledge. The goal of the test was to collect opinions and thoughts regarding the concept to be able to point out what was appreciated and what could be improved. The goal was also to collect data regarding how useful the solution was, by measuring success and time to complete the tasks.

Before the test started the participants were allowed to explore the prototype for a couple of minutes, so they felt comfortable interacting with the interface. This was a change that was made after the first test had been carried out, meaning that the first participant hadn't used the prototype before the test started. This might have affected the time it took to complete the task for the first participants, primarily the first task. The test with the first participants could therefore be viewed as a pilot test. However, since this deemed to mainly affect the result of the first task, the result from this participant is still included in the results.

The test was then carried out by giving the participants 5 tasks, which were presented verbally and written on a note that the participants could use as support during the test. The tasks were presented as scenarios based on situations that the annotators could face. Below follows a summary of the goal for each task:

1. Find the right class for a motorcycle.
2. Find which value should be chosen for a property when the correct value lies right between two values.
3. Find a correct value for a property regarding a specific type of vehicle.
4. Determine if a mannequin should be annotated or not.
5. Find where the extreme points should be placed when the object was only partly visible.

The participants were then allowed to ask questions before the test started. During the task, the participants were asked to think aloud, and the audio was recorded as well as notes regarding the

actions of the participant were taken. A task was completed when the participant had communicated to the facilitator what he/she thought was the right answer. After each task, the participant was asked to fill out a short questionnaire regarding difficulty, time consumption, and confidence in the correctness of the answer. After all tasks were completed, the participants were also asked to fill in some questions regarding the interface itself.

6.2 Results

Below the results of the delivery phase will be presented. This section is structured in a way which first presents the resulted structure followed by a presentation of the prototype and finally it will present the results of the test

6.2.1 Instruction headings

In this section, the heading structure which could be derived through looking at multiple GLs will be presented. It presents what was found to exist under these headings and thereby also what should exist under each heading. This structure was further the one used in the prototype and test.

6.2.1.1 Introduction

Contains some sort of introduction to the project. What could be found was that the GLs usually had some sort of task description, for example, to annotate vehicles, road marks, faces, seatbelts, etc. Some guidelines also contained information regarding the technology used to capture the data, which data format should be used, and abbreviations and vocabulary used in the GL. The just mentioned information and similar types of information were therefore deemed appropriate to collect in an introduction chapter. See example of how to use the introduction heading and suitable subheadings in Figure 12.

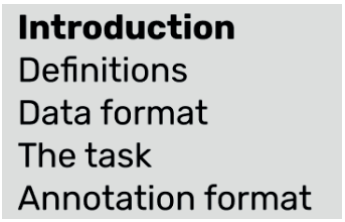


Figure 12 - Introduction heading examples

6.2.1.2 Annotation instruction

Under this heading some sort of instruction regarding how the actual annotation should be performed should be given. This could be information regarding where and how to mark an object, which objects should be annotated, which should be discarded, a proposed workflow, tolerances, etc. This information, which explains the annotation task in detail, but without explaining what constitutes a class, property, etc., should be put under this category, see example in Figure 13.

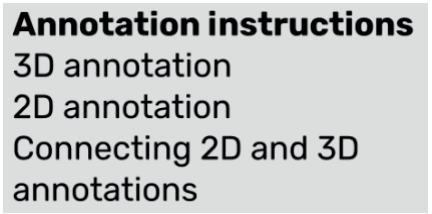


Figure 13 - Annotation instruction heading examples

6.2.1.3 Classes

This category collects information about the different classes that exists in the project. The types of classes and the number of them varied a lot in the different instructions. But they all had some sort of text explanation to what defined each class and usually some sort of picture examples about each class. The subheadings to the heading "Classes" should be the classes used in that project, see Figure 14 for an example.

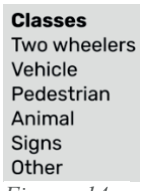


Figure 14 - Classes heading examples

6.2.1.4 Properties

After determining a class, the objects properties should be set. The properties are set by giving them specific values. For example, if an object has the class vehicle it is given a "type" property where the values can be car, truck, van etc. Usually, a class has several properties and some of these properties should be applied to all the classes, while some are specific to certain classes. The properties were usually explained in text with the different values which also were explain in text but often accompanied with pictures examples. Sometimes examples of when a criterion of a value wasn't met, or other types of edge cases were shown as well. In the structure created in this project it was decided that the properties which are general for all classes each should be its own subheading under the properties heading and the class specific properties be its own subheading but with the classes as sub-subheading, see Figure 15.

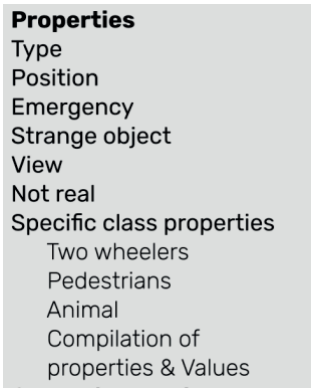


Figure 15 - Properties heading examples

6.2.1.5 General examples

Some of the GLs contained more general examples, usually of edge cases, or to show the entire annotation process. These were found difficult to put under a specific property or class, since they could explain several different of these, or explain something more general. These types of examples were collected under this category.

6.2.2 Figma-prototype

A complete description of the final prototype will be presented in the next chapter, 7 Final concept. However, in this section, a brief explanation of the results leading up to the final prototype will be given.

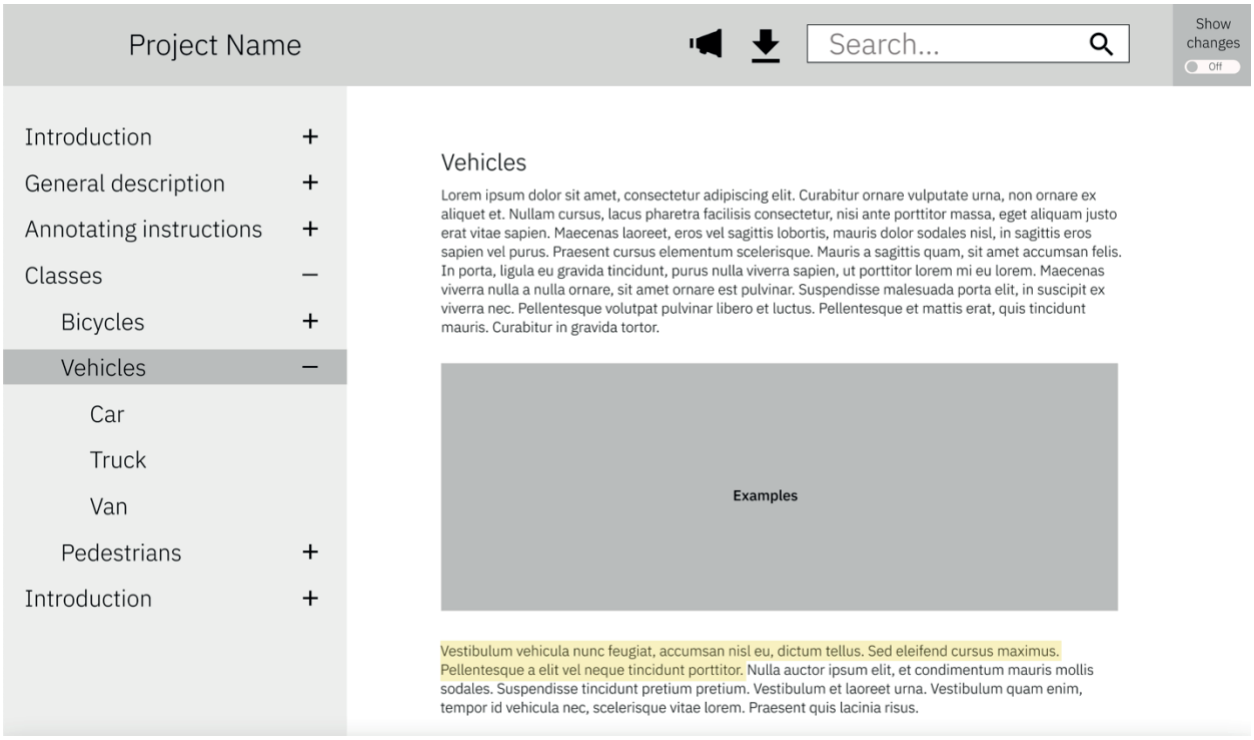


Figure 16 - First sketch of final prototype

In Figure 16 a first sketch of the final prototype can be seen. The goal of these sketches was to specify which elements should be included in the final concept and their placement. The main goal was to create them quickly to keep the focus on the basic structure and not go too much into detail, making the fidelity of these sketches rather low.

Much of the basic structure resembled the initial mock-up used in the evaluation. However, some changes were made. One change was the announcement centre, which came from discussions during the evaluation workshop regarding how changes could be communicated to ensure that the annotators understand the changes made. Another is the addition of a download button, to be able to download the entire instructions in a PDF. During the evaluation, comments were made regarding how different annotators might preferer to use the instructions. To make sure that all the annotators can use the instructions as they like, this seemed like a reasonable addition.

When the prototype was made the content was collected from an existing GL. This was done in order to test that a GL could be converted into the concept and to be able to use the prototype in the evaluation. However, since the GL is owned by the customer this meant that the prototype, including the instructions, cannot be presented here due to secrecy reasons. To be able to understand the concept made properties and classes were created for this report and the text was replaced with Lorem Ipsum.

6.2.3 Usability test

In this section the results from the usability test will be presented. These results didn't lead to any changes of the prototype, due to time constraints. They are presented here to provide an indication whether Eagle Eye is a promising solution or not and give input for future work.

The average and mean time it took as well as the number of participants who managed to complete the task is presented in in Table 11. On task 4 two numbers are presented because the task was divided into two subtasks. The first number states how many participants managed to complete both tasks, and the number in parentheses how many completed one of the tasks. As can be seen in the table task number 1 and 3 stands out in completion time and number of participants that correctly managed to solve the task. These two tasks were more straightforward forward where the participants were asked to find the right class in task 1 and a value for a property in task 3. The other two tasks required the participants to understand the instructions on a deeper level where they were asked to handle edge cases. Also, worth pointing out is that the average and mean time for task 1 differs quite a bit. This is due to the first participant, who took considerably longer time to complete this task. Since this participant hadn't used the platform before the test started this is an expected result. Therefore, the mean time for task 1 can be more meaningful to study than the average time.

Table 11 - Results from usability tests. Time measured in seconds.

Task	Average time	Mean time	Number of complete (max 4)
1 - Find the right class for a motorcycle.	61	12	4
2 - Find which value that should be chosen for a property when the correct value laid right between two values.	117,5	105	3
3 - Find a correct value for a property regarding a specific type of vehicle.	58	55	4
4 - Determine if a mannequin should be annotated or not.	195	165	2 (4)
5 - Find where the extreme points should be places when the object was only partly visible.	157,5	160	3

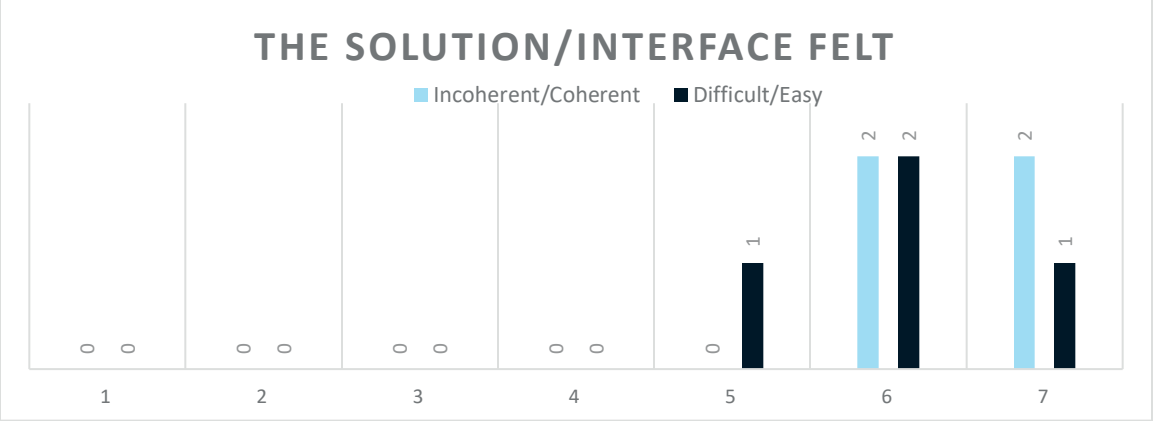
What can be said about these results is that the solution performed well when the task was to find information, but not quite as good when it came to comprehending the instructions on a deeper level. This is a result that could be expected since it was the first time the participants were

introduced to annotation instructions and had no introduction. To validate whether it was the novelty of the task or the new format itself that caused the task to be uncompleted or the duration a comparative test must be made. Such a test should be done with people of the same knowledge background and by giving them the same tasks but with the new versus the old solution as a tool, one can compare if it was the task in general that cause the problem or the solution that confuses. However, the conclusion that can be drawn from this is that the new solution shows potential when it comes to finding a specific type of information, which is an important aspect since this is a situation the annotators are commonly faced with during the production. However, how the solution performs when it comes to providing a deeper understanding of the instructions is something that would require further investigation.

Regarding the navigation in the prototype the participants were overall able to navigate with ease in the prototype. The only exception form this was the interaction required to access the actual instructions from the main menu. Once the participants understood how this navigation functioned, they had no problem interacting with it for the rest of the test. The conclusion can therefore be drawn that the interaction required to access the instruction form the menu isn't as intuitive as can be desired and might require further iteration.

After all the tasks were completed, the participants were asked to rate the interface. One of the questions asked was regarding how coherent the interface felt within itself. Since the test participants have nothing to compare to coherent reflects the consistency and compatibility as defined by Jordan (2002), see 2.4.2 Jordans principles of usable design. As can be seen in Table 12, the participants said that it felt coherent or very coherent. Creating a coherent solution can help users find information since distractions by a confusing interface/structure are removed. Whether it will be coherent between different projects, however, is something further testing needs to prove. Such test should use the structure given in 6.2.1 Instruction headings, use the same type of test with similar a task but use the GLs from different projects.

Table 12 - Table showing how difficult/easy and incoherent/coherent the participants found the solution. (1=very incoherent/difficult, 7=Very coherent/easy)



In Table 12 it can also be seen that the participants perceive the interface as somewhat easy to very easy. Since the information in the solution should be quick and easy to find it was deemed important that the interface felt easy to use, to make sure that the interface and its interaction weren't a distraction.

6.2.3.1 Conclusion usability test

Below a list is presented with the key insights from the usability test:

- Overall, the navigation was performed with ease. The only expectation is the first time navigating between the main menu and the instruction text.
- Specific information was easy to find in the prototype.
- Challenging to gain a deeper understanding of the instructions on the first use.
- The way which the prototype presents the instructions was appreciated and shows promise.

07.

FINAL CONCEPT

In this chapter, the final concept, Eagle Eye, will be presented in its final stage. The chapter will begin with a brief overview of the solution and will be followed by an in-depth description of the different elements. Furthermore, an account of how the proposed solution will affect the stakeholders will be given, as well as an explanation for how the concepts solve the posed requirements.

7.1 Overview

In Figure 17 an overview of the home screen can be seen with some of the main elements in the concepts. In the top of the screen the header is placed and is consistent in both appearance and placement throughout the interface. The header contains functionality related to search, download, version history and announcements. In the centre of the screen the main menu is placed, which consists of several menu items, the menu items can then be divided into heading, subheading and sub-subheading. It is via this menu the user can get an overview of the content of the instructions as well as access to them.

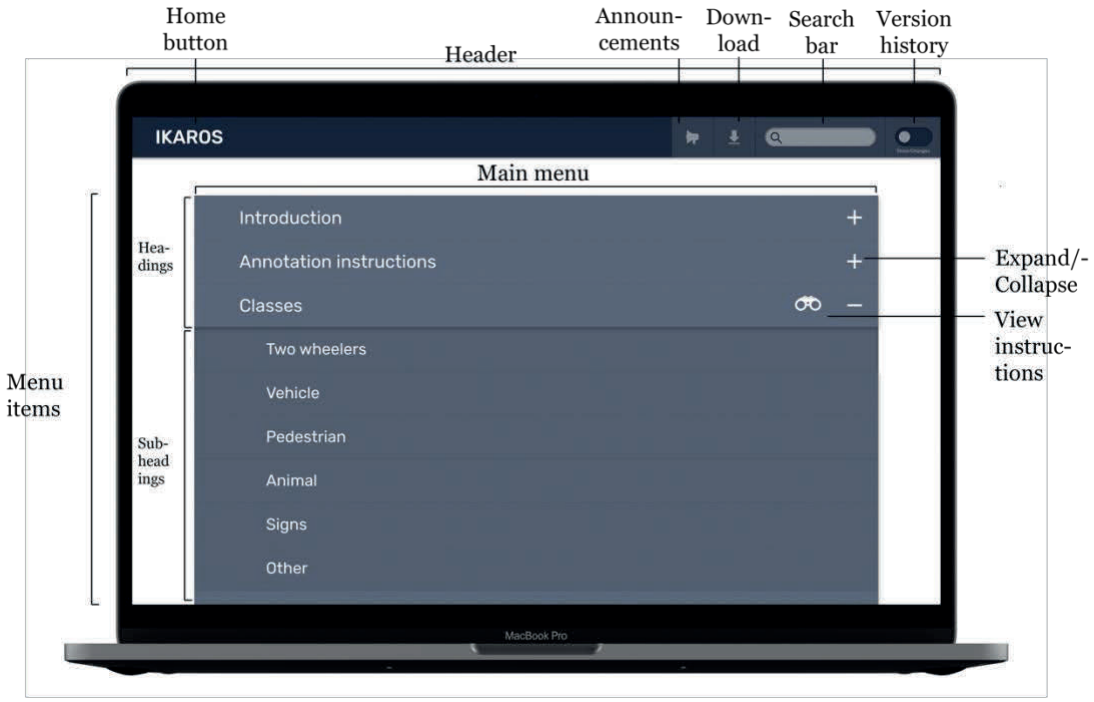


Figure 17 - Home screen with explanation of elements.

In Figure 18 the state of the interface when the user has entered the instructions can be seen, how one enters the instruction will be presented in 7.2.2 Home screen & main menu. The main menu has been replaced with a navigation menu, to the left of the UI and in the centre of the screen the actual instruction text can be seen.

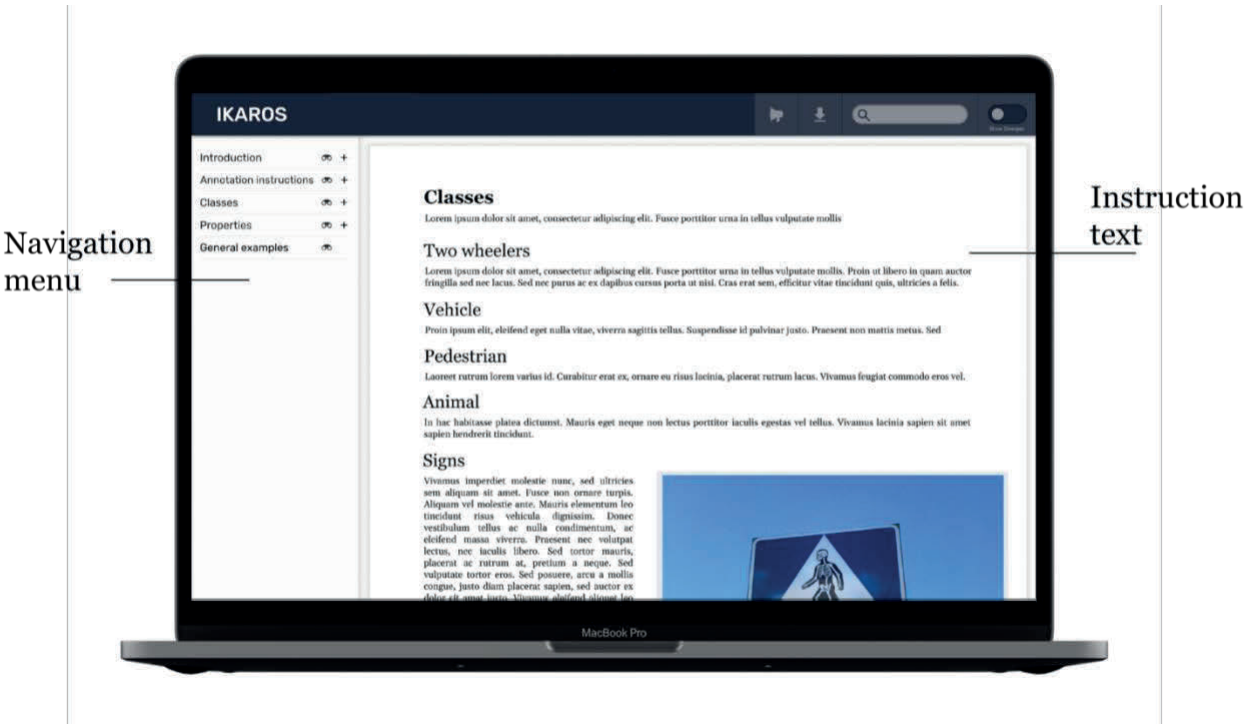


Figure 18 - Instructions view with text explanation.

7.2 Elements

In this section the different elements of the interface will be explained.

7.2.1 Header

The header is consistent throughout the interface, meaning it always looks the same and contains the same functionality. The header can be seen in Figure 19. To the far left the project name is displayed, in this case, “Ikaros”. This also functions as a home button, which brings the user back to the main menu. On the right side of the header, four different functionalities exist. First is a button that opens the announcement centre, indicated by a megaphone. Secondly, there is a download button, which allows the user to download the instructions in a PDF. Thirdly, a search bar is presented, which allows the user to search the entire instructions. Lastly, there’s a toggle switch that can be used to toggle on and off showing what changes has been made to the instructions.

Figure 19 - The header of the UI.

7.2.2 Home screen & main menu

In Figure 20, the view of the home screen and main menu can be seen. This is the first thing the users see when using the interface and the menu contains the overall structure of the instructions.

The first level of the menu is always expanded and shows the top headings of the instructions. The plus sign to the right of the heading indicates that the menu item is expandable.

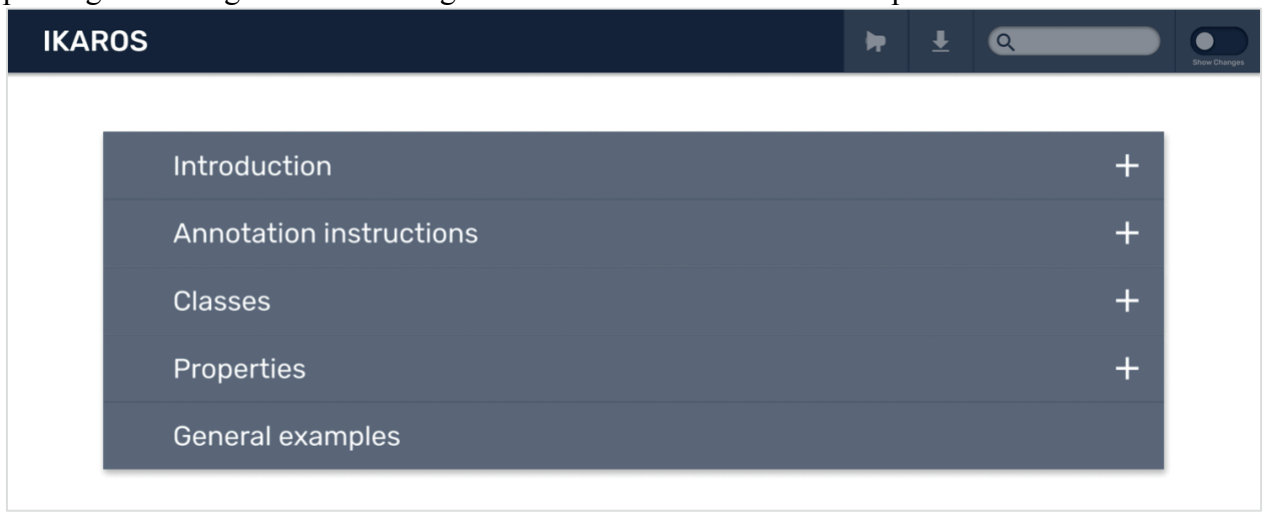


Figure 20 - The main menu, containing the overall structure of the instructions and is the first thing the user sees when opening the UI.

To expand a menu item the user can click anywhere on the heading button. While hovering the item a pair of binoculars will appear next to the expand symbol, as can be seen in Figure 21. If the binoculars are clicked, the user enters the instructions at that specific point.

Figure 21 - Two menu items, the upper one shows the hover state, which displays the binoculars.

When a menu item is expanded, it will appear as presented in Figure 22 where the item “2D annotation” is hovered. If there exists another level of headings the expand symbol will appear on this item. The maximum number of levels the menu can be expanded to is 3. 3 levels were deemed as shallow enough to find the way to the right heading yet deep enough to be subject-specific.

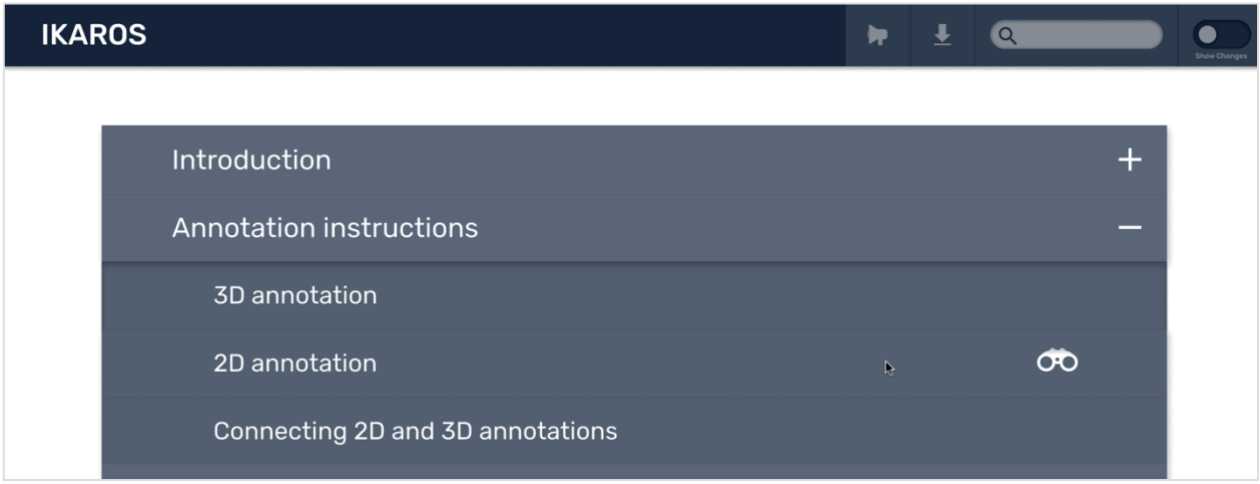


Figure 22 - One menu item expanded.

7.2.3 View of instructions

When the user clicks on the binoculars on a menu item, the corresponding instructions are displayed, as can be seen in Figure 23. Here all the necessary instructions for the annotators are displayed in the shape of text and images in a scrollable view. The headings in the text of course correspond to the headings presented in the main menu. To the left, the navigation menu is present, which will be described in the next section.

Figure 23 - View of the instructions with navigation bar to the left.

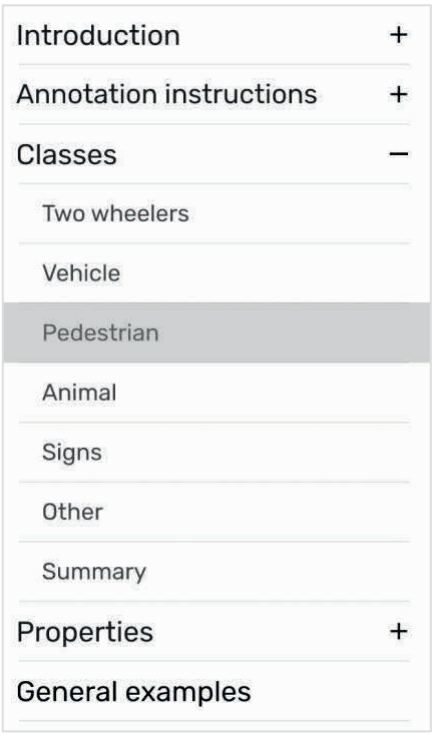


Figure 24 - Navigation menu where the user is currently located in "pedestrian.

7.2.4 Navigation menu

The navigation menu is located to the left of the screen when the user has the instruction view open. The content of the menu is the same as in the main menu but visually in a slightly different way due to the change of colours. The grey box in the menu, marked as "pedestrian" in Figure 24, indicated that the user is currently at that position in the instructions. This box follows as the user scrolls through the instructions, indicating the current position.

7.2.5 Sections

Each first level heading also constitutes a separate section in the instructions. When the user reaches the bottom of a section two buttons appear (given that there exist a previous and following section) which the user can use to navigate to these sections. An example of this can be seen in Figure 25.

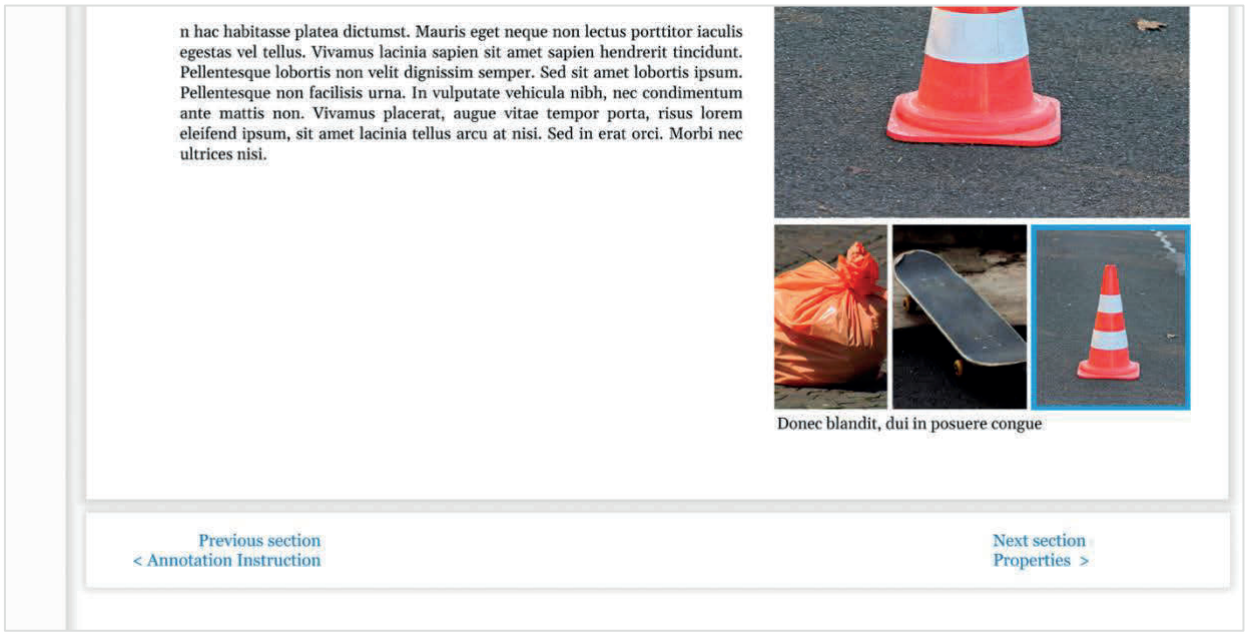


Figure 25 - Showing the bottom of the screen with buttons for previous and next section

At the end of each section these it is possible to add a summary of the information presented in the section, which can highlight important aspects in the instructions, see Figure 24.

7.2.6 Images

Pictures are displayed by showing one highlighted picture with a picture gallery below if there exist more pictures corresponding to the same section. The size of the box containing the gallery pictures can vary, depending on how the surrounding elements are placed. A minimum of one row of pictures in the gallery must be displayed and a maximum of 5 rows can be displayed at the same time, given that the gallery doesn't become bigger than the section or interferes with another picture.

Figure 26 - Example of how pictures are displayed, with a gallery of pictures below.

In Figure 26 a view of how pictures are presented is given. Here two rows in the gallery are displayed, with a third one hidden. The top edges of the images that are "hidden" help communicate to the user that the gallery is scrollable as well as the scroll bar to the right of the gallery. The bar also helps the user to understand where in the gallery he/she is positioned. In Figure 27 the same view as in Figure 26 is presented, but here with the gallery scrolled to the bottom. To communicate which picture is selected there exists a blue frame around the selected picture. Which can be seen in Figure 25 and Figure 26. The pictures can also be displayed in full screen. This is done by double clicking, either the miniature picture or the expanded picture.

Figure 27 - View of pictures with a gallery, with the gallery scrolled to the bottom.

7.2.7 Changes

Two different types of changes to the instructions can be made, these are communicated in two different ways. The different types of updates, as presented in 4.2.1.2 Problems with updates, can either be data changing updates or clarifying updates.

If the change is a data changing update the annotator will get a pop-up window in the Omara app, indicating that an update has been made. When the user opens Eagle Eye, a notification will be seen in the main menu, see Figure 28. The number inside the badge indicates the number of changes that have been made. As the user expands the menu items the badge will move to be located on the menu item corresponding to the change, this can be seen in Figure 29.

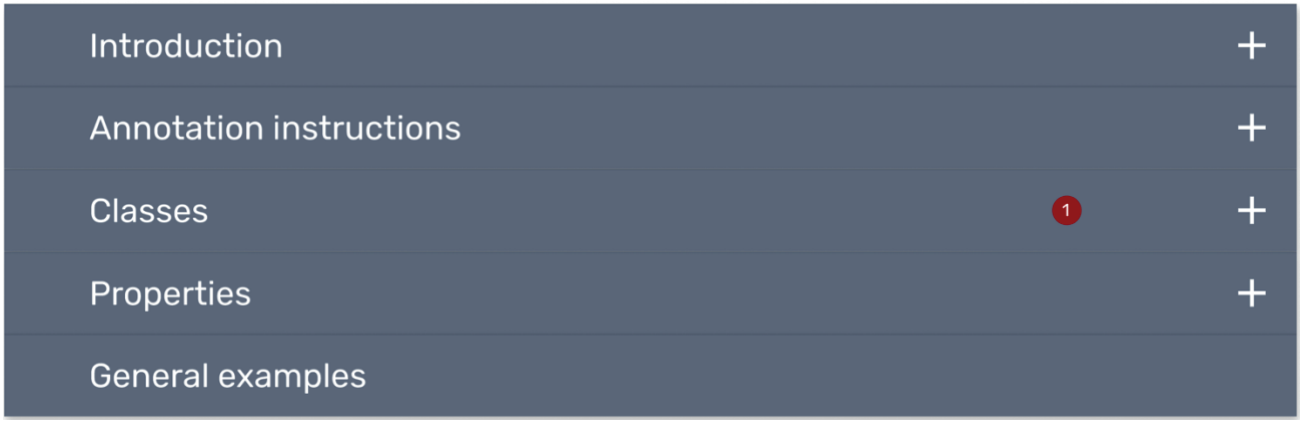


Figure 28 - Main menu with notification badge

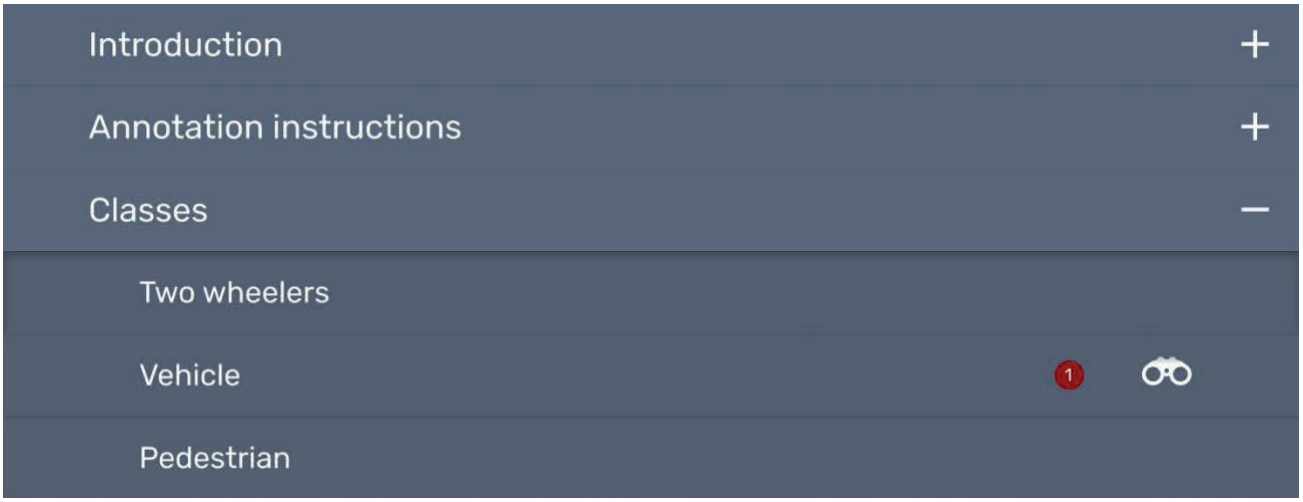


Figure 29 - Expanded menu item with notification while "vehicle" is hovered.

When the user has located the section in the instructions that contains the change, the change will be presented as can be seen in Figure 30. The red colour highlights where the change has been made and the crossed-out text indicates the part which has been removed (if information has been removed) and the red text indicates the new information. It will be displayed in this way until the user has spent a few seconds on that page without scrolling. Then until the next time the user opens the instructions, it will not be displayed as a change anymore. The reason for this is to ensure that the notification stays there until the user has read the change, so it doesn't disappear while quickly scrolling through that section. These changes are also communicated via announcements, see 7.2.9 Announcements.

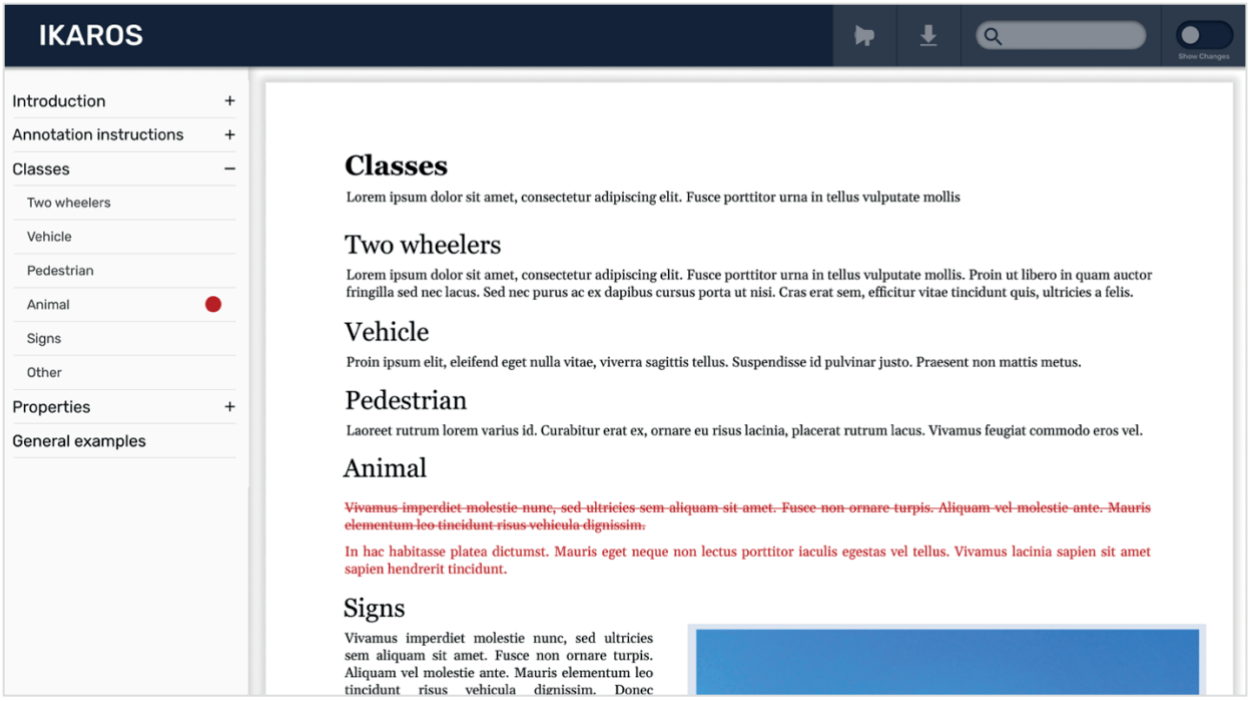


Figure 30 - View when a data changing update has been made.

The other type of updates, the clarification updates, doesn't follow the same presentation since they aren't as severe. Instead, they are only communicated via the announcements. One specific type of clarification update doesn't generate any type of notification to the user, nor ends up in the announcement, these are corrections of grammatical errors. Since they are clear mistakes and don't affect the user in any signific way they were deemed to be too unimportant to notify the user about.

7.2.8 Version history

The version history can be toggled on and off anywhere in the instructions, by clicking the button in the top right corner of the UI. When the version history is turned on all the changes made to the instructions become visible. In Figure 31, it can be seen how it looks with version history toggled on. Clarifying changes are green and data changing updates are red.

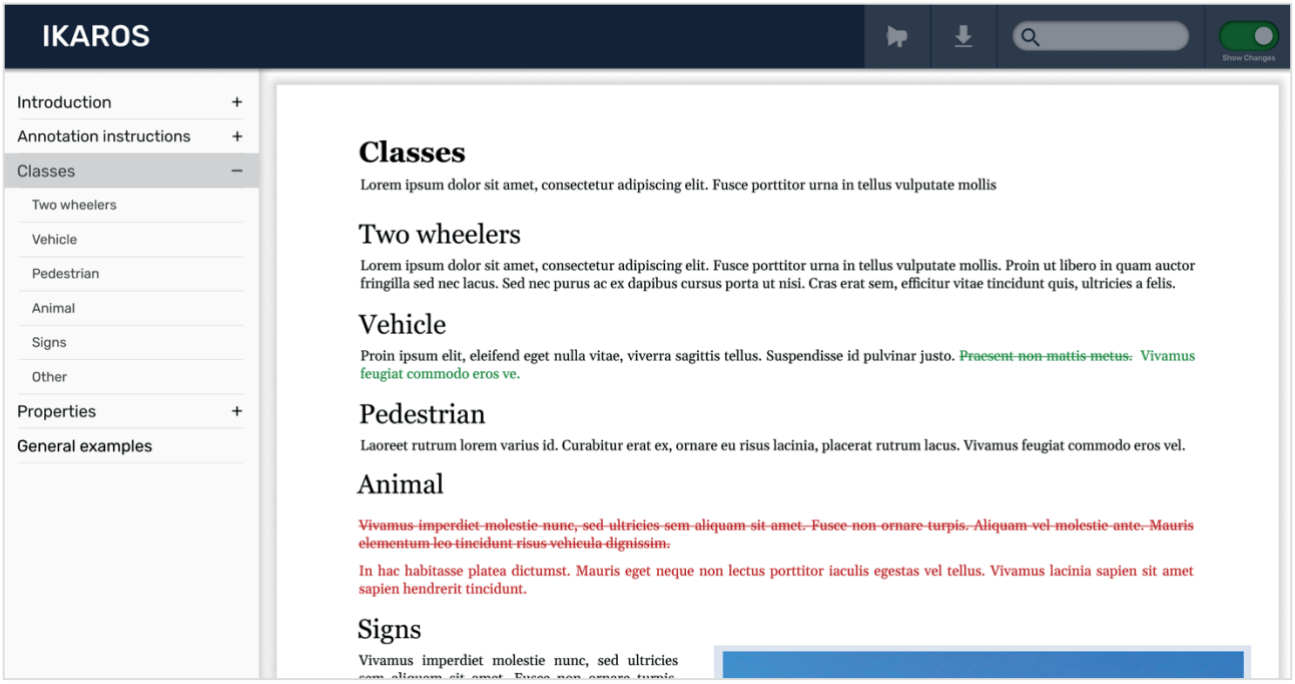


Figure 31 - View showing the instructions with version history toggled on. Green indicates clarifying updates and red data changing updates.

7.2.9 Announcements

When a change has been made, an announcement is published. The announcements communicates that a change has been made, as well as provide context about that change. The announcements can be accessed via the announcement centre, indicated by the megaphone in the header. If there is a new announcement it's indicated via a small red badge on the symbol. When this symbol is pressed the centre opens, as can be seen in Figure 32. Read announcements are blue and unread ones are indicated by the grey colour. Data changing updates are indicated by the red coloured heading. A limited amount of text to each announcement is presented in this view, to see the full information, the centre can be expanded by pressing "show all" at the bottom of the window.

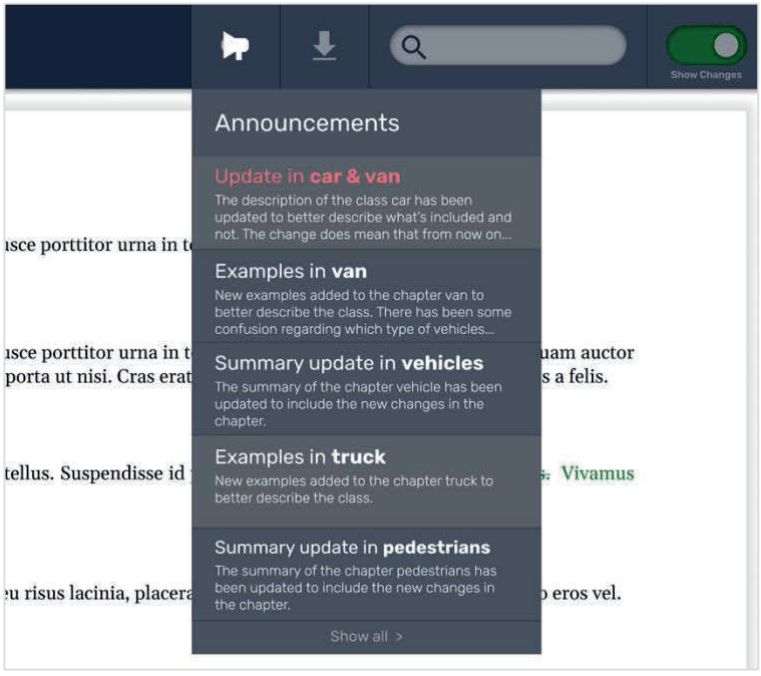


Figure 32 - Announcement centre opened.

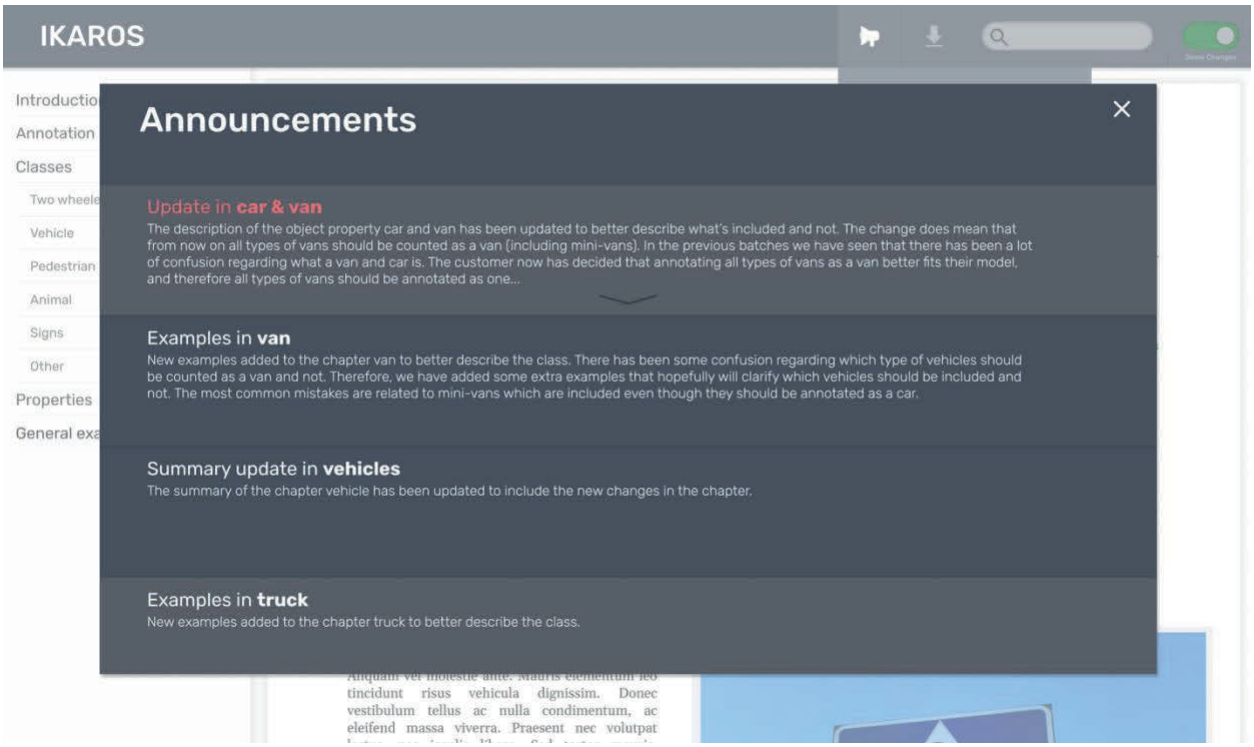


Figure 33 - expanded announcement centre.

In Figure 33 a view of the expanded view is presented, which is displayed via a pop-up window. Besides being able to see more information about each announcement the user can also scroll down in the window to see older notifications. Like the previous view, the amount of information displayed is limited to avoid each announcement taking up too much space. In the cases where the announcement contains more information than what is visible the announcement can be expanded by clicking on it, which can be seen in Figure 34. In the same figure the hover state for an announcement can be seen, in this case, the top announcement, which displays a button to the right stating “go to change”. When this button is pressed the announcement centre is closed and the instructions are opened at the position where the change has been made.

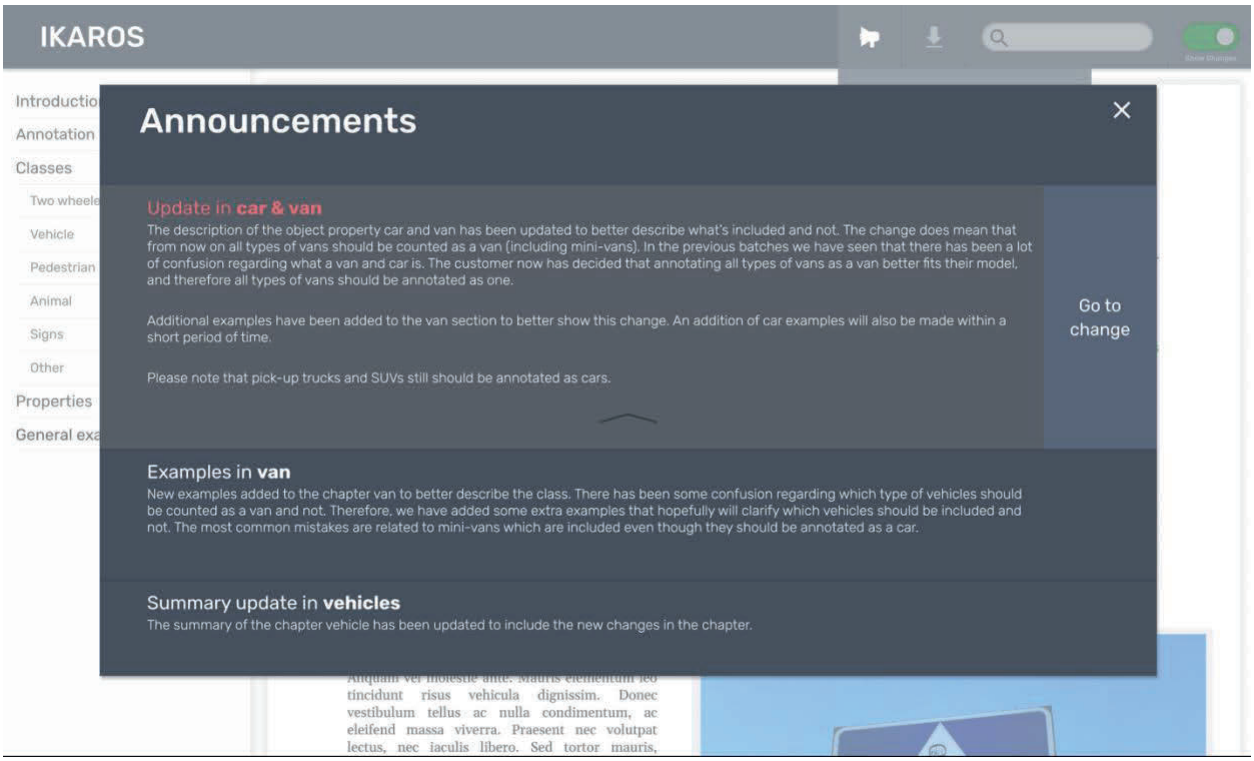


Figure 34 - View showing the full announcement text with the first item in hover state.

7.3 Application of Eagle Eye

In this section an explanation of how Eagle Eye can be applied and how it will affect the stakeholders will be given, with an emphasis of how the new solution will change the current situation.

7.3.1 Before production

Since the customers usually provide Omara with a GL they already have written, Eagle Eye was designed with this in mind. The idea is that the GL the customer supplies can be used and adapted to the concept without having to do any major changes to the instructions themselves. This is done by providing a flexible solution, in terms of what content (instruction text and images) and structure (headings) can be used in Eagle Eye. With the structure found in this project and the ease of updating the Eagle Eye, the instructions can easily be restructured in a way that presents the instructions similarly between all projects. Furthermore, it was found that the GLs provided sometimes had very scattered information without internal consistency. The structure would also help maintain internal consistency. Both these things will help *minimise the time and effort required for finding information* as well as providing *proximity*, two of the design guidelines mentioned by Osvalder and Ulfvengren (2015), see 2.2 Cognitive ergonomics

The fact that the original GL can be used reduces the effect the concept will have on the customers' way of working and thereby lowering the threshold for implementing the solution. However, adapting the instructions to Eagle Eye requires some work from Omara or possibly the customers

themselves. This would most likely increase the workload required by Omara and, given the roles existing today, the responsibility would most likely fall on the IE. However, the benefit of this is that no matter how different the GLs might look they still could be applied to Eagle Eye. This is an important aspect, given that the projects and their GLs look very different. This flexibility also makes it possible to arrange and rearrange the instructions logically.

With regards to the flexibility of the solution, there exists a balancing act between flexibility on one hand and consistency on the other. The solution needs to be flexible enough to fit a variety of different instructions while remaining consistent throughout the different projects so that the user can find the correct information where he/she expects it to be. This is balanced by allowing different content and subheading structures in the different projects while remaining a consistent interaction and layout throughout the different projects.

Before the production starts the IQM and/or IE have the opportunity to write a summary for each main section. This gives the annotator an overview of the content and the IQM/IE can point out important aspects of the instructions. These summaries can also be updated throughout the project to highlight topics and key points which have showed to be important or easily overlooked. Besides being able to highlight important topics the summaries also provide an opportunity for the annotator to repeat the information they have read previously in the instructions, increasing the possibility for the information to be encoded into the long-term memory. This means that the need to send out additional summaries, which is something that occurs today, is minimised.

7.3.2 During production

During the production, the goal is that the need to use Eagle Eye should be as small as possible, but when it's used it should be as effective and efficient as possible. Since the annotator should have gone through the instructions during the training process before the production starts, he/she probably has an idea of what type of information needs to be found and where this information exists. This means that when information needs to be found the user engages more in scanning than searching, based on Osvalder and Ulfvengren (2015) definition of the searching and scanning, meaning the user more systematically goes through the information with an idea of what this information looks like. Unlike scrolling through an entire PDF-file, the interface helps facilitate this process by dividing the information into different sections. In eagle eye the user can navigate to the section which contain the relevant information, which makes it easier to scan through each sub-section without being overloaded with information. The navigation menu, with an indication of the current position as well as previous and upcoming sections, further helps the user to understand the topic of the section and identify where the correct information might be. If the annotator knows more specifically what information he/she is searching for, the search bar can also be used. Since the results are presented in a structured list, unlike today's solution which only relies on the web browsers internal search function, the user gets a good overview of the available information and can choose the result which seems most probable to contain the desired information.

One relevant concept when designing the solution was the concept of re-usability. Since re-usability deals with the level of deteriorated performance a product has after the user haven't used it for a comparatively long period, it was deemed as an important aspect. Important because during the production, the instructions are mainly used when the annotator is unsure what to do. If the solution has good re-usability the situations when the user is required to use to solution should be kept to a minimum. The Eagle Eye is designed with Jordans principles of usable design in mind to enable this, see section 2.4.2.

The most requested improvement from the annotators was to include more examples. The limitations of a PDF-file make it difficult to present many images while remaining good visibility and overview of the pictures. How pictures are presented in Eagle Eye is designed to both provide a good overview of the available pictures while still being able to facilitate many images. By providing a scrollable box with thumbnails an overview of the available images is given, and with the possibility to expand each picture, the images can be viewed with good detail. Since the images are presented in a box the amount of space that the images claim can be reduced. This means that even if 100 pictures are presented in a gallery, the amount of screen estate the pictures claims will be no bigger than if it was 10 images, for example.

In today's solution, the PDF can be downloaded. This is a functionality deemed important to remain for two reasons. Firstly, this is the way the annotators use the instructions today, and by being able to download the instruction as a PDF the threshold to start using the solution might be lowered. Secondly, the customers might be interested in accessing the instructions as PDFs to be able to communicate the instructions internally or with other annotation companies.

7.3.3 Updates

In today's solution, updates can only be made by releasing an entirely new version of the GL. This means that smaller updates or clarifications might be delayed if they aren't deemed important enough to single-headedly trigger an update of the PDF. To make sure the annotators still get the information it might be communicated in other ways, such as via documents the IQM/BPOQM writes. With Eagle Eye, the struggle of updating the instructions is minimised since changes can be made specifically to the relevant section without having to release an entirely new version.

Since updates can be made continuously the solutions need to make users aware of the change at the appropriate time and with a level of intrusiveness that reflects the severity of the change. The information in a clarifying update isn't crucial for the annotator to take part of instantly and is thus only communicated via announcements. On the other hand, the annotator must read the information regarding data changing updates right away, therefore these are communicated in a more intrusive way by triggering a pop-up, red notifications badges in the main menu, and highlights of the changed text in red colour.

To help the user to understand why the change has been made and how it affects the annotations, more information regarding the change can be provided in the announcement centre. In the

announcement, the IE, or possibly IQM, can give information regarding the change, which today probably would have been provided separately, for example in a newsletter. The announcement centre would also function as a history of previous changes. So, if the user wants to go back and study changes made at an earlier state it is possible.

When a data changing update has been made both the old and the new information is displayed, in red colour and with the old information crossed over, to emphasise the change. This isn't something that occurs when a clarifying change has been made, to differentiate the two different update types and minimise the risk of the interface becoming cluttered with highlighted text. This was also something that wasn't deemed to be necessary since information regarding the change is provided in the announcement. However, if the user wants to view how the instructions have changed, it is possible to turn on version history and, in that way, see all the changes made.

Finally, the persons issuing the updates and announcements should be limited to ensure that the updates are valid and consistent with the rest of the instructions. This is a responsibility that probably should fall on the IE and possibly the IQM.

7.4 Connection to user study and requirements

In this section, the final concept will be described and analysed from the problems defined in 4.2.1 Current situation, as well as how the solution meets the requirements states in 4.2.3 Requirements.

7.4.1 Communication of instructions

That the new solution will fulfil the main- and subfunction, "Solution shall communicate the annotation task" and "Solution should contain instruction to fulfil the task", is vital and arguably the solution does so. The main function is enabled by providing a solution that communicates instructions in any shape or form, and the subfunction is dependent on the customer as long as the new solution allows for instructions to be written. Furthermore, the text and figures which the customers provide will be displayed in Eagle Eye, as presented earlier in this chapter, the requirement "Text, including figure texts, must be readable" will be fulfilled.

The annotators receive information from multiple sources today. Since the PDF is a static document and thereby cumbersome to update, it is hard to avoid that managers on different levels feel the need to complement the exciting instructions. Since the proposed solution is a web-based platform that has the potential to enables quick and easy updates, it can in turn help reduce the number of extra documents, fulfilling the requirement "Should minimise the amounts of documents". Furthermore, it makes the instructions easily sharable, fulfilling the requirement "Should be digitally sharable", in addition being web-based also enables the requirement "Should have the ability to include images, GIFs and videos". The addition of summaries at the end of each chapter further contributes to keeping the number of documents low. Here the IE or IQM can add extra clarifications and summarise the most important points from the chapter. Since this is

possible to update during the production process and address questions that have been raised, this would further reduce the need for extra documents.

It can therefore be argued that Eagle Eye helps streamline the communication by reducing the need for separate documents to be sent out to the annotators by the IQM/BPOQM. This information can instead be included in updates/announcements or summaries. This also increases the transparency, since it easily allows all the involved stakeholders to view the information which is communicated to the annotators. All the communication won't be possible to communicate via Eagle Eye, but the concept can help reduce the number of separate documents needed to successfully communicate instructions. Thereby decreasing the risk of miscommunication as well as increasing the transparency of the communicated instructions.

7.4.2 Updates

The PDF is a static document implying that once it has been sent out is cumbersome to update. The only way the GL can be updated is by releasing an updated version of the document. With Eagle Eye, updates can quickly and easily be sent out to all the relevant annotators. Once the update has been released the new version replaces the old, reducing the risk of confusion with different versions and instead provides a more seamless experience. This also means that only the relevant part of the information is updated without having to send out a new version of the entire document. This indicates that the requirement "Should be easy to update/edit", will be fulfilled. The announcement centre makes it possible to back-track updates that have been made previously as well as provide an overview of the updates and context to these. And if the user wants to study the exact change the version history can be used. Furthermore, the announcement centre and the notifications of Eagle Eye makes the user highly aware of updates and changes, fulfilling the requirement "Should make the user aware of changes" and "Should communicate when a change has been made and the content of that change"

7.4.3 Inconsistency

Since each GL is adapted into Eagle Eye a much higher level of consistency can be maintained than within the previous solution, while still using the same information. This applies to both aesthetic factors as well as structural ones. Since all the instructions are presented in the same UI, they will naturally have some similarities in terms of appearance. To keep a consistent presentation, as mentioned by Osvalder and Ulfvengren (2015) in 2.2 Cognitive ergonomics the colour scheme used in the concept is inspired by the colours used in the Omara app. Other colours used in the UI, are red, which indicates data changing updates, and green, which indicates that something is turned on, in this case, the version history. The green colour is also used to indicate clarifying updates which are only shown if - show changes - is turned on. The red and green colour are also chosen to keep a consistent presentation, match between system and the real world & consistency and standards which also are design principles mentioned by Osvalder and Ulfvengren (2015). Other functions such as the home buttons placed at the top left, the usage of symbols and

icons, and placements and structures of menus are also designed with those design principles in mind.

Since the GL is restructured when it is converted into the platform it is also possible to restructure it in a way that is consistent between different projects. Of course, it is impossible to keep everything consistent in every single aspect, but with some restructuring and consistent way of communication a significant improvement can be made.

7.4.4 Understandability

In the user study issues regarding the understandability of the instructions were raised. Since it still is the customers who will continue to write the instructions, Eagle Eye might not solve all the problems related to this topic. But with a combination of the proposed functionality (mainly the possibility to restructure instructions, update the instructions, write summaries, and create announcements), the level of understandability hopefully can be raised. However, when it comes to the language used in the instructions, it is harder to guarantee that the requirement related to this will be solved based on the presented solution since it still will be the customers who write the instructions. Hence the requirement "Should only use simple English" is not guaranteed to be fulfilled. With the same arguments also the requirement "Should be easy to comprehend regardless of technical background" cannot be guaranteed by the solution since it is more dependent on how the customer writes the instruction.

7.4.5 Speed and quality

By dividing the information into chunks in the shape of chapters related to headings, and division within the chapters into sub and sub-subheadings, the information should be easier to take in with the lowered risk of information overload. By prioritisation of information, the user should know which heading to search for information within and which subheading to study, to find the sought information. Hopefully, this solution helps with the absorption of information and enhances the ability to remember information. If it does, it enhances the speed and quality just by enhancing the learning process. However, if it does not increase the ability to remember, the previously mentioned aspects regarding chucks and headings should minimise the time spent to find information, which would increase the speed and remain the quality.

Furthermore, the search function should also possess qualities that help to enhance the speed and quality of the annotation. This by making it faster and easier to find the desired information. This could also reflect positively on the quality since the annotators might become more prone to use the solution and find the desired information. This means that the requirements "Solution should help enhance quality whilst remaining speed", "Solution should help minimise annotation time", "Solution should be effective and efficient to use during production", and "Should be able to quickly find the desired content" should be fulfilled.

7.4.6 Compilation

In Table 13 a comparison of the old and new solution is presented. The comparison is a compilation of the problems and solutions presented earlier in this section, based on the problem areas. In the above sections, some of the requirements rated 3 – Important and all requirements rated 4 - Very Important or above are presented. In Table 14 the remaining requirements are stated with an explanation of if and how Eagle Eye enables that requirement.

Table 13 - A summary of how the current solution and eagle eye compare in relation to the identified problem areas.

Problem area	**Current solution**	**Eagle Eye**
Communication of instructions	• Complementary instructions communicated via other sources that the GL (summaries, newsletter, chats, verbal communication) • Possible several intermediate steps (IE-IQM-BPOQM-Annotator) • Separate communication channels	• Combine summaries in platform • Easy to updates to reduce need of extra document • Coherent and transparent source of information. • The need for communication in separate channels will remain in some cases.
Problems with updates	• New PDF-file on each update • Might be hard to understand the changes made	• Eliminate the problem with different versions of the GL • Announcements • Toggle version history • Possibly pop-up
Inconsistency	• Each customer writes and structure the GL on their own • No GL looks like the other	• Coherent UI • Possibility to restructure the instructions
Understandability	• Inconsistent structure and language • Cumbersome to update • Multiple sources of information	• Restructure instructions • Updates • Summaries • Announcements
Speed and quality	• One long text • High risk of information overload	• Chunks of information • Clearer division of sections • Search function

Table 14 - Remaining requirements

Requirement	**Weight**	**Eagle Eye**
Solution should help customer to formulate what they want annotated/produced	2	The solution has a generated structure which helps structure and indicates what building blocks should exist in the solution.
Solution should increase probability to cover most scenarios	3	By being easy to update, scenarios and edge cases which occurs that aren't already explained can be easily added. The coverage of scenarios and edge cases isn't guaranteed from the start but can be added and explained
Solution should increase probability to cover most edge cases	3	

		along the way. It is hard to predict edge cases but what can be done is to help with instructions when they do occur.
Customer should not need to have insights in the annotation tool	1	This is maintained since the customer will not need to gain any more knowledge regarding Eagle Eye versus the old solution.
Customer should not need to be in contact with many different people at Omara	2	
Should contain relevant examples for the data set being annotated	3	This is requirements that relies on the customer to fulfil. Eagle Eye however enables the possibility to upload many examples yet having them in good proximity to the instruction needed to be exemplified. Tables can as easily be added as well.
Should contain many examples	3	
Lists and tables should be preferred over texts descriptions	2	
When referring to something it should be in direct proximity	3	
Should express instructions rather than delivery	3	This is a requirement that relies on the customer and IE, this solution will neither enable nor disable this.
Should have a consistent structure	3	With the structure created it should have consistent structure between the different projects as well as within each project.
Should have concise information regarding the annotation task and reduce amount of assumptions needed	3	There is nothing in this solution that enables or disables this. It should however help the annotators console the instructions due to the more effective way to reach information rather than making assumptions. If there on the contrary is a lack of instructions, assumptions might not be avoided.
Length of figure texts should be limited	3	Relies on the customer. Nothing this solution will enable or disable.
Should provide high resolution pictures	2	
Should allow annotators to feel safe to ask any question	2	This solution will neither enable nor disable the way of communication. Most likely it will remain the way it is today. It will however become easier to clarify things to the annotators of a project via updates with this solution which should streamline the communication in at least one direction.
Should allow questions to be raised in an effective and efficient way to and between annotator, BPOQM, IQM, IE and customer	2	
Should have a high level of learnability	3	This means that the user should feel satisfaction using the solution again after having used it once. It should, based upon using design principles, enable learnability and encourage usage.
Should use prioritisation of information and be concise	3	Eagle Eye only displays the amount of information needed to navigate to the next stage etc. until the stage where the instructions searched for is situated. Whether the instructions are concise relies how the instructions are formulated.
Should encourage to use the solution rather than consulting BPOQM	3	Hopefully the structure of the Eagle Eye will make it more likely for the annotators to console the instructions rather than the BPOQM in those cases where clear information exists.

08.

DISCUSSION

In this chapter the methods used, the results of these, the process of the project, and future recommendations related to the project will be discussed.

8.1 User study

To be able to answer the first goal of this project, how the GL is used today and the problems with this solution, a user study was carried out. In this section, the methods, and results of the user study will be discussed.

8.1.1 Selection of methods and participants

In the user study, interviews and a survey were the main methods used to gather data. Interviews were used to understand the general problems with the current situation and the different roles involved in creating and using the instructions and their relation to each other. The interviews also gave insights regarding the IE, IQM, and BPOQM opinions and thoughts regarding the GL and the current situation. The survey was used to gain a deeper insight into the annotator's situation and their opinions regarding the GL.

One of the biggest challenges in the project, especially related to the user study, was the inaccessibility of the annotators. Since the annotators were located in other countries, and not directly employed by Omara, it was difficult to conduct interviews with them. Partly due to the language barrier, which probably would require an interpreter to overcome, and partly because the interview situation might make it hard to receive honest answers. People at Omara who had conducted interviews with annotators stated that honest answers could be hard to receive since the annotators were inclined to give positive answers regarding Omara. It was also stated that the company for which the annotator works not seldom wanted a manager present at the interviews, which could further hinder the annotator to feel comfortable, open, and honest during the interviews. For the same reasons, it was also difficult to conduct observations of the annotators, especially direct observations. Observations could have brought valuable insights regarding how the annotators use the GLs in their day-to-day work. Given these constraints, these insights instead had to be based on the survey which was sent out to the annotators and on the interviews with people who work with the annotators. The survey was a good solution to still gain the opinions of the annotators. However, a survey will never give as qualitative information as an interview. A survey on the other hand generates quantitative information which also is highly valuable. Although a mix of the two would have been preferred.

The customers of Omara are stakeholders who haven't been interviewed. There existed an ambition to conduct such an interview, but it wasn't found possible to arrange. Attempts were made but without success. Since the annotators are the main stakeholders for this project, interviews with the customers were not deemed crucial for the project, it could however have contributed to a solution that could be more beneficial for all stakeholders. To compensate for the lack of contact with customers, and to gain their insights on the matter, pre-recorded interviews with two customers were viewed and analysed.

8.1.2 Interviews

All the interviews were conducted via video calls, due to different geographical placements and/or Covid-19. This made it possible to interview persons which otherwise would be inaccessible, and it made recording and transcribing the interviews easy. The downside of these types of interviews is that it might be harder to create a good connection between the interviewer and interviewee. Since the interviewee's day-to-day job include a lot of remote work, this wasn't found to be a big issue. However, in some of the interviews, especially those where it also existed a language barrier it became noticeable.

The results from the interviews with the IEs were deemed highly valuable and trustworthy. They helped define the problem and identify the entire process from constructing the GL to the delivery of annotations. The interviews with the IQMs gave mixed results where some gave more valuable results than others. It was found that some of the IQMs were rather new whilst some had been working as IQMs for a long time, which resulted in data that representing different degree of experience. The same results were found with the BPOQMs where one was experienced and one rather novice. Regarding the BPOQMs it could have been beneficial to conduct one or a few more interviews to ensure that the opinions gathered were general and not specific for the interviewed persons. One of the most interesting finding with both the IQMs and the BPOQMs was that summaries were written where one IQM and one BPOQM stated that they supply summaries to the annotators. Overall, the interviews generated a lot of valuable insights regarding the problem areas as well as an understanding for the annotator's situation. The recorded interviews with the customers generated a limited amount of valuable date, since questions from this project's point of view could not be asked. It however gave an insight that customers are interested in new solutions if they generate annotations of higher quality.

8.1.3 Survey

The survey which was sent out to the annotators had a very high response rate, 228 of around 250 annotators responded. This means that almost all the annotators who are currently hired by Omara responded to the survey. Given that the survey was the only method that had the annotators as participants, this is an important aspect.

The majority of the questions were closed-ended which is to be preferred in surveys. However, since interviews with the annotators weren't executed, some of the questions in the survey had to be posed as open-ended questions. This led to a big variety in the answers, which in turn made them hard to analyse. If options had been provided in some of these questions more meaningful quantitative data could have been extracted but alternatives would also influence the annotators' thoughts. Since these questions were more open for interpretation by the respondents, some of the answers were found to be irrelevant to the question or hard to interpret. However, the data extracted from these questions are by no means useless, general trends could be identified and the answers helped to paint a picture of the current situation. One example of this is when the annotators were asked about what could be improved with the GL. Since this was posed as an open question a big

variety of responses was given. Most of the responses were related to the GL, however, some of the responses concerned other aspects, such as the Omara app. This meant that each response had to be interpreted and determined if it was related to the posed question or not. If more knowledge about what areas of improvement the annotators saw with the GL had been acquired, the questions could have been given as a closed question, thus making the data easier to interpret and reducing the risk of misunderstandings. This could however lead the annotators to think in a certain way or force the annotators to pick an alternative and not formulate their own answers.

The high response rate implied that the results from the survey could be used to conclude the annotators' perception regarding the situation and GL and the survey was therefore deemed reliable. However, since this survey was anonymous and did not require any type of log-in it couldn't be controlled whether an annotator answered the survey several times to "correct" their answer. Some of the answers were concerningly similar which sparked this thought. But in conclusion, the number of answers should make the possible misleading answers neglible.

8.2 Concepts and solution

In this section the focus will be on the second goal of the project, to find a solution that helps accelerate the annotators' learning process of instructions. The discussion will be based on the produced solution, as well as other potential solutions.

8.2.1 Focusing on the GL

Early in the project, the scope was narrowed down to focus on the GL and how it is communicated to the annotators. The reason for this is that the GL is the main communication method and the single most important document regarding the annotation instruction. It is what ties together the customers' demands with what the annotators produce. However, what can be discussed is if the GL is the only aspect that can be improved or if another approach could have been taken to accelerate the annotators' learning process. One such approach could have been to instead of focusing on the GL itself, focus on the training process before the production starts. How the annotators are presented with the instruction, the training tasks, and examination are aspects that could have been investigated, if the scope had been defined differently.

8.2.2 Customer and annotator

One reoccurring conflict in the project is the balance between the two stakeholder groups customers and annotators. Since the main goal of the project has been to improve the annotators' learning process, the needs of the annotators have been the most important to meet. However, the customers could not be completely ignored in the project since its they who order and funds the process. It is the customers who write the instructions and if the solution realistically should be implementable, the needs of the customer can't be discarded.

8.2.3 The concept generation

One can always discuss whether the ideation stage has been fully examined. During this project there was many methods used to ideate but other methods could of course have been used. The phase started wide and general and sub ideas and solutions were generated through brainstorming. The ideas were then conceptualised by using a morphological matrix in a specific way. The way it was used in this project does not generate as many ideas as the matrix is designed to produce. It was used in a way which hand-picked ideas to create concepts. By doing so not all possible solutions were examined but the time spent on the method was reduced. In the aspect of generating valuable concepts whilst minimising time this was deemed defendable. Not only due to the fact that it reduces time by not creating all possible concept, it also saves time with screening the concepts generated, a process which is more time-consuming than actual creating the concepts. Therefore, it was deemed defendable to use the method in this way and using the time into refining the concepts generated. One thing that can be more thoroughly discussed is the part regarding how the solution enables communication was fully discarded. The reason for this being discarded was due to, no matter which concept was chosen, any type of communication method could be added in a later stage. Therefore, a decision to completely discard the internal communication within the concept was made and said to be a further area of improvement. By doing this focus could be kept on producing a concept that communicates the instructions in an effective and efficient way.

8.2.4 Concept evaluation, refining, and screening

The evaluation methods used, six thinking hats, pugh-matrix, and evaluation workshop all worked as refining evaluations as well as screening methods. Through six thinking hats many perspectives of the concept could be discussed. This resulted in screening of concepts that were deemed insufficient, whilst some ideas of sub-solutions, deemed superior, was incorporated into other concepts. The six thinking hats method also gives a human approach to the evaluation, and aspect such as ones “gut feeling” can be considered. On the other hand, the pugh-matrix is focused mainly on estimations and requirements, and is less monitored by human feelings. A downside to the pugh-matrix is that some of the less important requirements still can give a rather large impact on the result. This was kept in mind while screening. For example, The PDF did not receive a good number but was still deemed as good to keep due to it being easy to implement and not being too impactful on the annotators' way of working. The six thinking hats combined with the pugh-matrix gave a good estimation of which concepts would be more efficient and effective in meeting the requirements, whilst still considering the concepts holistically, which the human gut-feeling contributed to. These methods resulted in a screening that left three concepts. These three concepts were then presented to four people at Omara in an evaluation workshop. The evaluation workshop gave great insight into how implementable the concepts were to the company as well as which concepts were most realistic to give the outcome desired e.g., to help the annotators learn, understand, and use the instructions.

8.2.5 The final concept and prototype

The main objective of the proposed solution is to accelerate the annotators learning process of annotation instructions. Eagle Eye aims to solve this by facilitate the location of information, improving the way which the information is presented, and reducing the amount of information sources and thereby also making the information available to as many as possible.

To make the information easier to locate the information is divided into sections and introduces a navigation menu, which allows the user to easily navigate between the different sections, as opposed to scrolling through a PDF-file. The division into section also helps to improve the way which information is presented and minimise the risk of overflowing the annotator with information, as the amount of information presented in each view is limited. Another way which the presentation of information is improved, is the way which examples are presented. The gallery allows for many examples to be presented, while remaining proximity between the examples and the corresponding information in the text. However, one potential risk with this exists. If too many examples are presented, the annotators might stop to rely on the actual definitions in the instructions and instead solely rely on the examples. For example, if the examples were to contain images of 100 different types of cars, and the annotator is suddenly faced with a type of car which is not included in the examples, there is a risk of the annotator becoming unsure whether to include that car or not.

By introducing summaries in Eagle Eye, the number of information sources can be reduced. Instead of having to send out separate documents containing this information, the summary can be added to the rest of the instructions. This has the potential to make it easier for the annotators to keep track of the instructions, as they are gathered in one place. Furthermore, by introducing summaries in the concept it's easier to ensure that all the annotators receive the same information, instead of the annotators receiving different information depending on who is their BPOQM/IQM. Although worth pointing out is that Eagle Eye doesn't remove the need for information and instructions to be communicated via other channels. Questions will still appear, and it might not be meaningful to include every piece of information in the instructions, as it could risk cluttering the instructions. However, by making the instructions easy to update, more frequently asked questions and identified pain points, the instructions themselves or the summaries can be updated to clarify these issues. Making the answers available to more annotators than the ones that posed the questions.

Related to the learning of annotation instructions, Eagle Eye aims to accelerate this process mainly by making the information quicker to find when the user needs the information. This is done by the structure, navigation and search bar. The goal is that the annotator should easily and quickly be able to find the desired information, thus facilitate the learning process. This approach was chosen since it has the potential to increase the quality of the annotations with a concrete solution that Omara can implement. Another approach to the problem could have been to focus more on the initial learning of the instructions and the training process before the production starts.

However, this would increase the dependency of stakeholders further down the production chain, both regarding user studies as well as the implementation of the solution. This is by no means impossible, but given the level of uncertainties that would arise if the project had turned in that direction the focus was kept on the main communication of the instruction, the GL.

This solution would probably increase the workload for Omara, as the solution needs to be built, and the GL translated into the platform when a new project starts. Hopefully the trade-off in potentially increased quality of the annotations would make Omara deem this increased workload worth it. Depending on how the concept is implemented, the possibility exists to allow the customers themselves to enter the instructions in Eagle Eye. This would decrease the workload for Omara, however it could risk leading to an inconsistency in the structure of the instructions, as Omara would lose this control.

8.3 Further development

In this section possible further development will be discussed as well as any unanswered questions that needs further investigation.

8.3.1 Editor tool

The scope of this project has been to develop a solution that accelerates the learning process for the annotators and the solution has been developed from this perspective. Therefore, one aspect of the solution has not been addressed, namely the editor view (i.e., how the instructions are entered into Eagle Eye and maintained). This includes the process of entering the text and images into the platform, writing summaries and producing updates. To be able to execute this some sort of editor tool needs to be developed. The stakeholders and requirements for such a tool were deemed to be outside the scope of the project and not investigated fully enough to produce such a tool.

The user of such a tool would be someone at Omara, for example IE or another role with similar responsibilities, and/or possible the customers themselves. Depending on how an editor tool is designed it would be possible that the customers can produce their instructions directly in the editor tool. This would reduce the workload for Omara and make the entire process more seamless. This would require that there's an incentive for the customers to do this work, possibly a difference in price for the different services.

8.3.2 Communication

The problems with the direct communication between the IE, IQM, BPOQM and annotators is something that has been brought up in the user study and is also reflected in some of the requirements. The main issue identified was that the communication method used can vary between the different roles and from annotator to annotator. In turn this led to the annotator receiving information in several different channels. The solution aims to partly solve this, since Eagle Eye is easier to update and makes it possible to include more information than the GL, in the shape of summaries. However, the need for direct communication, for example when the

annotator encounters a problem and needs to ask a question to their manager, will remain. A solution to this could be to create a communication tool, which can be used when the annotators encounter a problem and is accessible for all involved stakeholders.

When investigating this problem, it became apparent there's a lot of stakeholders involved, and issues started to arise which were deemed to lay outside the scope of the project. In early iterations of the concept there existed a chat functionality in the platform, but this was later removed. It became apparent that this would only create new problems rather than solving them. A chat function in the platform would create another communication channel for the annotators, without making the other ones obsolete, thus only escalating the problem it was aiming to solve.

To effectively creating a new communication tool a holistic analysis of the current situation needs to be made, in order to fully understand the requirements for such a solution. This comprehensive study was not deemed to be possible to make given the frame of this project.

8.3.3 Further testing

In order to gain a deeper understanding of how the new solution compares to the current one, more comprehensive user testing needs to be executed. The usability test showed that the solution has potential to solve the stated problems, but in order to determine how well these problems are solved and if it is an improvement on today's solution, a more profound study needs to be carried out.

Two different types of studies can be imagined. Firstly, a more in-depth usability study where the entire interface is tested an evaluated. In the preformed test, the users got to test some basic interactions and locating information. Other aspects that could be tested includes: Search functionality, announcements, updates, image interaction, summaries, and version history. Secondly, a long-term test of the learning process of the instruction could be performed. In order to ensure that the solution has a positive effect on the learning process, an evaluation over a period longer than the traditional lab-based usability studies is required. One possible approach to such an evaluation could be to provide one test group with the current solution and one with Eagle Eye and let them use then in the same annotation project for a longer time period. Possible metrics are the overall time to complete the annotation tasks, the quality, and "test" where the annotators are asked about the content of the instructions.

8.3.4 Application during training

The concept in its current state mainly focusses on the use during the production process. However, the concepts potential to aid the training process is something that could be further investigated. One possible approach to this could be to make the training more structured, perhaps with additions of gamification elements. For example, quizzes could be deployed during certain stages in the instructions, like at the end of each chapter. This could help the annotators to reflect upon the information they just consumed and make them more observant while going through the instructions. This could help the initial learning of the instructions to become more structured and

ensure that all the annotators receive the same training and information. This isn't something that necessarily have to be exclusive for the training process, but something that could reoccur during the production process to refresh the annotators memory.

8.4 Ethics

Since the annotators are spread around the world, they come from a variety of different cultures and language prerequisites. In the beginning of the project some discussions were held related to the difficulty of the language in the GLs and how it effects the understandability of the instructions. However, it became apparent that this was a hard task to investigate. Questions related to language skills were posed in the survey to the annotators, but it was hard to draw any conclusions from this as it was based on the annotator's own estimations of their English skills. To be able to draw any meaningful conclusions related to language skills some sort of test would've been required. This, however, wasn't executed partly due to the risk of such a test becoming too intrusive, and partly because this was deemed difficult to solve within the scope of this project, as it mainly would fall under the realm of the customers way of writing the instructions. Related to cultural differences, there might exist aspects of the solution that could be interpreted differently than what was indented. One example of this is the use of colours. In Eagle Eye the red colour is used to emphasise the importance of crucial changes. While this might make sense in some cultures, it might not make it in every, where the red colour might have other connotations. The use of symbols, such as the megaphone for the announcement centre, and the binoculars for the view-instructions-button are examples of other elements in the concept where the interpretation might differ depending on the user's cultural background. These aspects are important to keep in mind in future testing of the solution, to ensure that any misunderstanding doesn't occur based on the user's cultural background.

Another important aspect when discussing ethic is accessibility. However, generally in this project, accessibility hasn't played a crucial part. One example of this is when information is communicated via the colours red and green, which might make the interface harder to use for a colour-blind person. The reason for this limited attention the accessibility aspects is the requirements which the annotation task has on the ones executing it. It is a task that fully depends on the visual sense and sometimes colours are used in a way which would make it hard for a colour-blind person to perform the tasks.

Finally, one external factor which has influenced this project is the war in Ukraine. When the project started some of Omara's operations were based in Ukraine. Some of the interviews held in this project was with participants from Ukraine at a time before the invasion, but still in a very unstable situation. As the events which would follow where hard to foresee at that point in time, the interviews where at the time deemed justifiable to conduct and provided a lot of valuable insights.

09.

CONCLUSION

In this chapter the conclusion to the project can be found.

END

The goal of this project has been to investigate and suggest how instructions can be communicated to annotators in order to accelerate the learning of the instructions. This has been done by analysing the current situation, collecting user needs and requirements, and further by using the insights to create a concept that aims to make the learning easier and more efficient. Since the main communication of the instructions is through the GL, the GL became the main focus of this project.

Even though the annotators were found to be rather content with the GLs, areas of improvements could still be identified. The user study showed that the most frequently requested improvement from the annotators was more examples. Furthermore, the GL is a static document which in turn leads to obstacles when updates need to be made, since a new version of the GL needs to be sent out. It was also quickly identified that different GLs differs in structure and appearance, as the different GLs have different authors. Besides the GL it was also found that the annotators receive other complementary instructions from IQMs and BPOQMs, in the shape of summaries, newsletters, and messages via internal communication channels. Furthermore, the IQMs also witnessed that the amount of information could be overwhelming for the annotators and generate a feeling of ”information overload”.

To help overcome these obstacles the interactive tool Eagle Eye is presented. By breaking down the instructions into sections and presenting them in different views, the amount of information presented can be limited, thus decreasing the risk of overloading the user with information. This also helps to increase the consistency, since all the instructions will share the same UI and the possibility exist to structure the instructions in a more consistent manner. Furthermore, Eagle Eye will make updates easier to implement, since it is possible to update the specific section in question. To help the annotator keep track of updates and create a deeper understanding regarding why they’ve been made, the announcement centre is introduced. This, in combination with the possibility to add summaries in Eagle Eye, will help reduce the number of sources the annotators receive information from. Finally, a larger quantity of examples can be presented in a more effective manner compared to a PDF.

The usability test indicated that the solution showed promise, especially regarding the interaction and locating information of a simpler nature. However, the long-term effects the solution will have on the ability to learn the instructions is something that will need further investigation and more comprehensive studies.

10. BIBLIOGRAPHY

American Psychological Association. (n.d.). Human-machine system. In *APA Dictionary of Psychology.* Retrieved May 11, 2022, from https://dictionary.apa.org/human-machine-system

de Bono, E. (2000). *Six thinking hats*. (Rev. ed.). Penguin Books.

Benyon, D. (2010). Envisionment. In *Designing Interactive Systems*. (pp. 177-197). Addison-Wesley.

Bligård, L-O. (2015). ACD[3]: *Utvecklingsprocessen ur ett människa- maskinperspektiv*. Chalmers tekniska högskola, Göteborg. Hämtad från http://www.acd3.se/assets/files/ACD3%20-%20Utvecklingsprocessen%20ur%20ett%20manniska-maskinperspektiv%202.1.pdf.

British Library. (n.d.). *What is SWOT analysis?* Retrieved May 17, 2022, from https://www.bl.uk/business-and-ip-centre/articles/what-is-swot-analysis

Cambridge Dictionary. (n.d.) Cognition. In *Cambridge dictionary.* Retrieved May 24, 2022, from https://dictionary.cambridge.org/dictionary/english/cognition

Dam, R.F., & Teo, T.S. (n.d.). *Affinity Diagrams – Learn How to Cluster and Bundle Ideas and Facts*. Interaction Design Foundation. Retrieved March 5, 2020, from https://www.interaction-design.org/literature/article/affinity-diagrams-learn-how-to-cluster-and-bundle-ideas-and-facts

Design Council. (2015, March 17). *What is the framework for innovation? Design Council's evolved Double Diamond*. Retrieved March 17, 2022, from https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond.

Google. (n.d). *Solution Sketch.* Design sprint kit. Retrieved May 10, 2022, from https://designsprintkit.withgoogle.com/methodology/phase3-sketch/solution-sketch.

Haupt, C., & Huber, A. B. (2008) How axons see their way--axonal guidance in the visual system. *FBL*, 13(8), 3136–3149. https://doi.org/10.2741/2915

ISO. (2018). *Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts* (ISO Standard No. 9241-11). Retrieved from https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

Jordan, P. W. (2002). *An introduction to usability*. Boca Raton. CRC Press.

Lim, Y.-K., Stolterman, E. & Tenenberg, J. (2008). The Anatomy of Prototypes: Prototypes as Filters, Prototypes as Manifestations of Design Ideas. *ACM Transactions on Computer-Human Interaction*, 15(2), 1-27, http://doi.acm.org/10.1145/ 1375761.1375762

Moran, K. (2019, December 1). *Usability Testing 101*. Nielsen Norman Group. https://www.nngroup.com/articles/usability-testing-101/

Muller, G. (2009). *Concept selection; Theory and Practice* [white paper]. Buskerud University Collage. https://www.gaudisite.nl/SESGwhitePaperNovember2009.pdf

Nielsen. J. (2012, January 3). *Usability 101: Introduction to Usability.* Nielsen Norman Group. https://www.nngroup.com/articles/usability-101-introduction-to-usability/

Nielsen, J. (2020, November 15). *10 Usability Heuristics for User Interface Design*. Nielsen Norman Group. https://www.nngroup.com/articles/ten-usability-heuristics/

Osvalder, A.-L., Rose, L., & Karlsson, S. (2015). Methods. In M. Bohgard, S. Karlsson, E. Lovén, A.-L- Osvalder, L. Rose & P. Ulfvengren (Eds.), *Work and technology on human terms* (2nd ed. pp. 473-576). Prevent.

Osvalder, A.-L., & Ulfvengren, P. (2015). Human-machine systems. In M. Bohgard, S. Karlsson, E. Lovén, A.-L- Osvalder, L. Rose & P. Ulfvengren (Eds.), *Work and technology on human terms* (2nd ed. pp. 349-435). Prevent.

Preece, J., Rogers, Y., & Sharp, H. (2015). Interaction design – Beyond human-computer interaction. (4th edition). John Wiley Sons Ltd.

Rousay, V (2021, Jan 21). Bottom-up processing. *Simply Psychology*. www.simplypsychology.org/bottom-up-processing.html

Smith, A. D., & Kelly, A. (2015). Cognitive processes. *The encyclopedia of adulthood and aging*, 1-4. https://doi.org/10.1002/9781118521373.wbeaa213

Ulrich, K., Eppinger, S., Yang, M. (2020). *Product Design and Development* (7th ed.). New York: McGraw-Hill Education.

UXPin. (n.d.). *Prototype vs. Wireframe vs. Mockup – What Are the Differences?* Retrieved May 17, 2022, from https://www.uxpin.com/studio/blog/prototypes-wireframes-mockup-difference/

Van Boeijen, A., & Daalhuizen, J. (Eds.). (2010). *Delft Design Guide*. Faculteit Industrieel Ontwerpen.

Wadhwani, P. & Loomba, S. (2022). Data Annotation Tools Market. Global market insights. Retrieved May 20, 2022, from https://www.gminsights.com/industry-analysis/data-annotation-tools-market

Wheelwright, S. C. & Clark, K. B. (1992). *Revolutionizing Product Development:* Quantum Leaps in Speed, Efficiency and Quality. The Free Press

Wikberg Nilsson, Å., Eriksson, Å., & Törlind, P. (2015). *Design: Process och metod* (1:3. edition). Studentlitteratur AB.

A.
APPENDIX

A.1 Personas

Bhavika
Annotator

Country India
Works for Gooseberry
Worked for 3 months
Native Tamil
English Above avrage

About Bhavika
Bhavika is an extremely thorough person. She is meticulous in everything she does and has therefore the highest quality rating in the entire company on annotations even though she only has worked for less than three months! There is no need to wonder if she will execute something with high results but her achilles heel is time management. She asks a lot of questions so that she always feels 100% secure with what she's doing and would never just "take a guess" and often replies with "why" to explanations so that she can get the full picture. Knowledge is key for Bhavika and perfectionist is something she's been called her whole life, and she secretly likes it. She has only worked at gooseberry for a little less than three months but has already made a strong impression.

How would you describe your job?
FUN! I love my job! And I'm good at it too! My job is to do annotations and we are currently marking signs. We did the onboarding exam like five weeks ago and I passed with excellence. Many of my colleagues are sloppy but fast. I am thorough at all times so that we know that everything produced by me is of highest quality.

What is your relation to the managers?
It's a love-hate relationship haha! I do a great job which they appreciate, I am rather trusted. But they can get frustrated with all the questions, I would never just do something if I'm not sure about what to do, so I ask my QM alot when I get stuck but sometimes I even feel as though they are just guessing. They are on me that I need to pick up my speed and I'm working on that but I would not tolerate it if the speed impacts my quality. I have the highest respect for Omara and the IQMs! They are my role models, I want to become IQM myself.

What would you say is the goal in your work?
Being the best! No, but I guess it depends on who you're asking. My goal is to become IQM but I think my "roles" goal is to deliver high quality annotations within a certain timeframe.

What is the most frustrating part in your job?
That I feel as though I don't get enough information. I need all the information to do a good job. Also sometimes with updates, they can be so contradictory and confusing. Like, you just learnt it being one way and now you need to "relearn" it all and sometimes it is like "do it the total opposite".

How and for what purpose do you use the guideline?
I use it for everything. I know it by heart. I read it so that i know exact quotes from it haha! No but honestly they are very important, it contains all the info so if you don't know it by heart you can't be as correct as you should and also if you need to look everything up all the time that is time consuming.

What is your impression of the guideline?
I like them! I mean they can contain more examples and be more comprehensive but i like them. I like reading so for me these guidelines are good!

Joshua
Annotator

Country Philippines
Works for Elderberries
Worked for 3 years
Native Tagalog
English Avrage

About Joshua
Joshua is 25 years old and has worked as an annotator for about 3 years in total, and for the past 2 year he has worked for Omara. When Joshua finds something that he likes to do or is passionate about, he commits fully to that purpose. However, when doing something which Joshua doesn't find as meaningful his motivation sometimes declines. He isn't a lazy or sloppy person, but sometimes Joshua's quest for shortcuts can go a bit too far. This sometimes leads to a lack of quality in his work, as long as he doesn't find the work meaningful and rewarding.

How would you describe your job?
A normal day I would say starts with going through the discord channel and email to see if I have received any new information or feedback. I then go to the annotation platform and start to annotate, this is where I spend most of my time. Sometimes questions appear when I annotate. The first thing I do is scroll through the summaries or guideline, if I can't find the answer there I send a message on discord to my QM or ask him directly if he is nearby. Usually he can answer right away, but sometimes he might need to check something in the guideline or ask the IQM.

What is your relation to the managers?
My relationship with my manager is overall good I would say. He is really helpful and can answer most of my questions. Sometimes I feel that some of the information could be communicated better, like if there's a change or something. Like, first I should do things in one way, and then all of a sudden I should do things in another way. This can sometimes create a bit of frustration if it's not communicated properly.

What would you say is the goal in your work?
To annotate as many tasks that I can, while still staying above the quality goals. The speed is usually not a problem for me, but sometimes I can receive some quality feedback. This is usually objects that I've missed, or that I've accidentally annotated objects that shouldn't be annotated.

What is the most frustrating part in your job?
I would say that the most frustrating part is that it is so hard to find information. When I encounter a problem or an uncertainty it can be really hard to know how I can quickly find the answer. I like the summaries but they don't cover everything, so then I might have to open and look in the guideline anyway. This takes a lot of time and the correct information is hard to find. So sometimes I just trust my gut feeling instead and annotate the way I think it should be done. I have annotated for 3 years now, so I think I have a quite good picture of how things usually are done.

How and for what purpose do you use the guideline?
In the beginning of a project I use the guideline quite a lot, it's used in the training and I look stuff up the first days of annotation when I'm still new to the project. But I use it less and less as I learn more. When I use the guideline I scroll through it and try to find the information that I need or use the search bar in the browser. Although I prefer to look in the summaries from the manager, because they are shorter and more easy to find information in.

What is your impression of the guideline?
The guidelines are quite big and hard to find information in. I don't like to use and look stuff up in the guideline because it takes too long to find and that takes time away from my speed goal. Instead I usually trust my gut feeling instead of double checking the guideline, if I were to double check everything it would take a very long time. If I have done something wrong I learn this from the feedback documents that I receive from my manager.

Datu
BPOQM

Country Philippines
Works for Cloudberry
Worked for 6 months
English Fairly good

About Datu
Datu is a driven, 29 year old, ex-annotator that has shown off his skill which led to a promotion to QM 6 months ago. He is very proud of his work and wants to inspire his team to be the best team, so that he can show off how good he is at his job as QM. He likes direct communications with his annotators, which has been hard for him during the times the Philippines has been in lock-down. He can come off as rather strict but is actually very soft and tentative when he is not under pressure. He has all the good qualities to be a good leader but is sometimes too goal oriented and can then tend to forget the well-being of his annotators. When he starts a new project he is very soft and good at teaching his annotators the task and how to think. He uses very pedagogical approaches but when the deadline is closing in he can get a bit stressed and forget his softer side.

How would you describe your job?
Well, I am the quality manager of 15 annotators currently. Our task is to annotate data for Omara, using their platform and deliver to them data that reach a certain quality. We are usually sitting in the same room whilst annotating but with all that has happened with covid and uncertainties we've been working from home alot. My job in specific is to be like a role model for the annotators. I train them on the task, answer their questions, and manage them. If I can't answer the questions I do have regular contact with the IQM. Most questions are resolved through us but if there is a more fundamental issue we also talk to the team at Omara.

What is your relation to the annotators?
I have a very close contact with the annotators! I tell them when they've done a good job and when they fall behind on quality or speed. I try to keep a good balance in my management style but if they don't pass the onboarding their standards aren't good enough and have to be swapped. I was an annotator before becoming an BPOQM so i know what they are going through and what problems they can be facing.

What is your relation to Omara?
Well... indirectly they are my employer, but i don't have much contact with them directly. Usually I have contact with the IQM but in some cases we all sit in a meeting so I have some contact with them.

What would you say is the goal in your work?
Ah, well I take a lot of pride in what I do so for me, having a team that delivers good quality to Omara and within a certain time frame makes me feel fulfilled. I of course want everyone to feel good in their work environment, but the main goal for me and my team is to be the best team at Cloudberry, an opinion I think is shared among my team members.

What is the most frustrating part in your job?
Haha, when I feel like no one is listening. It's like you can give an annotator the same feedback on a problem but yet the exact same problem reoccurs over and over again. That is very frustrating. Also with the questions, I feel as though they are sometimes too lazy to check the guideline or something and ask obvious questions. That slowed down whilst we went remote then I felt as though I got too few questions.

How and for what purpose do you use the guideline?
For all purposes almost! I mean it is my job description and that's all I get so I have to basically use it for everything. But after a while you memorize things, this is something which is both good and bad since it makes it more efficient but it could risk generating more errors.

Olena
IQM

Country Ukraine
Works for Omara
Worked for 3 years
English Fluent

About Olena
Olena is 32 years old and has been working for Omara almost as long as the company has existed. She started working as an annotator for Omara around 3 years ago and she has then advanced to the role as an IQM, which she has had for about 1,5 years now. As a person Olena is structured, ambitious and to the point. She isn't afraid to tell people her opinion and she doesn't beat around the bush. Her ways of communicating might come off as harsh sometimes, but that isn't because she is a bad person, it's just the way she is. When someone has done a job that doesn't meet the expectations she lets the person know, but she also makes sure to tell people when they have done a good job as well. If there's a problem she makes sure to go to the bottom with it. There's is alway a reason for why things aren't working and she will make sure it's found.

How would you describe your job?
I'm managing the BPOQMs and annotators, and is the link between the annotators and Omara. When we get a new project I go through the guideline and comment on things which are unclear and could be improved. I'm also the one that approves the examination tasks when a new annotator starts a project. I answer questions the BPOQMs might have regarding specific annotations or questions related to the guideline and try to explain them. If I don't know the answer I can always ask the in-house expert. During the production I measure the quality and speed, which is provided to the BPOQMs and annotators, as well as the customers. If it's necessary I can also make corrections and golden annotations.

What is your relation to the annotators?
I usually don't have direct contact with the annotators. Sometimes we have meetings with all the annotators in a project, for example when we start a new project or when it has been a big update in the guideline. But usually the communication goes through the BPOQM.

What is your relation to Omara?
I would say that I have a close relationship with Omara, even though I don't sit at their office I am still employed by them. I have much contact with several roles at Omara, such as the in-house expert.

What would you say is the goal in your work?
The goal of my job is to make sure that the data that is being delivered is of sufficient quality by making sure that the annotators reach their quality and speed goals. I do this by helping them understanding the guidelines, doing corrections and golden annotation, giving feedback to their annotators about speed/quality and pointing out errors which have been made. Sometimes, I also create summaries of the guideline to help the annotators understand them. This can be done based on questions which the annotators might have had, but also on my previous experiences from other projects.

What is the most frustrating part in your job?
That the annotators keep making the same mistakes over and over again. We try to send them feedback on what they have done wrong, but the same mistakes keep occurring. Maybe they don't read the feedback, or maybe they have forgotten it. But it is annoying when you see the same mistakes beeing made. Sometimes I don't think the annotators read the guideline that closely.

How and for what purpose do you use the guideline?
I use it to make summaries and communicate the instructions to the BPOQMs and annotators. It is used as a basis for our work and when something is unclear it is always the guideline that is first consulted.

A.2 Morphological matrix

Main Communication method	VR	AR	Webpage, Interactive digital document, "roadsign style"	Everything embedded in the tool. Adapts info to where you are in the platform	Voice	PDF	AI "chatt bot"	training process	Tables	PGE - Personal GL expert	a person sits and looks through all the pictures and categorizes them then there are ...	
Examples communicted	tutorials	popup with "examples"	know what kind of example should be communicated	pictures	Gif	in the platfomr	separate screen with exampels	"lexicon"	searchable image bank	link	instagram account "how others have done"	
Make user aware of changes	pop up	redtext, "lingon berries"	irritating sound	quiz	"disabled annotation platform"	time frame you have to look at text	extra punishable in qualitycheck	mail	slack	read receipt	versionhistory	onboarding
Find desired information	searchbar	"FAQ"	hashtags	"maybe this was really what you were looking for"	"Hey siri"	clearer paragraph breaks	index	simpler communication method	reward system "ask a friend" friend who answers will be rewarded	Presents only the information you need	Filtering function	
Should allow questions to be raised in an effective and efficient way	in the platform, cutting tools so you can take a print screen send to the desired person	chatt	chattbot	FAQ	forums - open anonymous questions, others annotate can answer as well	hotline	support errands	somments with @s	live call with screens			

A.3 Pugh-matrix

Category	Nr	Requirement	Weight	Current solution	Solution 1	Solution 2	Solution 7	Solution 10	Solution 11	Solution 12	Solution 13
Main Function	1	Solution shall communicate the annotation task	Must	0							
Sub Function	2	Solution should contain instruction to fullfill the task	5	0	0	0	0	0	0	0	0
Quality	3	Solution should help enhance quality whilst remaining speed	4	0	1	1	1	1	1	1	1
	4	Solution should help customer to formulate what they want annotated/produced	2	0	0	0	0	0	0	0	0
	5	Solution should increase probability to cover most scenarios	3	0	1	1	0	1	1	0	1
	6	Solution should increase probability to cover most edge cases	3	0	1	1	0	1	1	0	1
	7	Solution should help minimize annotation time	4	0	1	1	1	1	1	1	0
	8	Solution should be effective and efficient to use during production	4	0	1	1	1	1	1	0	1
Customer	9	Customer should not need to have insights in the annotation tool	1	0	0	0	0	0	0	0	0
	10	Customer should not need to be in contact with many different people at annotell	2	0	-1	-1	0	-1	-1	0	-1
Content	11	Should contain relevant examples for the data set being annotated	3	0	1	1	0	1	1	0	1
	12	Should contain many examples	3	0	0	0	0	0	0	0	0
	13	Should make the user aware of changes	4	0	1	1	0	1	1	1	1
	14	Should only use simple english	4	0	0	0	0	0	0	0	0
	15	Should express instructions rather than delivery	3	0	1	1	0	1	1	0	1
	16	Should have a consistent structure	3	0	1	1	1	1	1	0	1
	17	Should have concise information regarding the annotation task and reduce amount of assumptions needed	3	0	1	1	1	1	1	0	1
	18	Lists and tables should be preferred over texts descriptions	2	0	0	1	1	0	1	0	1
	19	When referring to something it should be in direct proximity	3	0	1	1	1	1	1	0	1
	20	Text, including figure texts, must be readable	5	0	0	0	0	0	0	0	0
	21	Length of figure texts should be limited	3	0	1	1	0	1	1	0	1
Technicality	22	Should be easy to update/edit	4	0	1	1	0	1	1	0	1
	23	Should be digitally sharable	5	0	0	0	0	0	0	0	0
	24	Should be easy to comprehend regardless of technical background	4	0	0	0	0	0	0	0	0
	25	Should be able to quickly find the desired content	3	0	1	1	1	1	1	0	1
	26	Should provide high resolution pictures	2	0	1	1	0	1	1	0	1
	27	Should communicate when a change has been made and the content of that change	3	0	1	1	0	1	1	1	1
Communication	28	Should allow annotators to feel safe to ask any question	2	0	1	1	1	1	1	0	1
	29	Should allow questions to be raised in an effective and efficient way to and between annotator, BPOQM, AQM, PE and customer	2	0	1	0	0	1	1	0	0
Other	30	Should have a high level of learnability	3	0	1	1	1	1	1	1	1
	31	Should use prioritization of information and be concise	3	0	1	1	0	1	1	0	1
	32	Should minimize the amounts of documents	3	0	1	1	1	1	1	1	1
Wishes	33	Should have the ability to include images, GIFs and videos	4	0	1	1	0	1	1	0	1
	34	Should encourage to use the solution rather than consulting BPOQM	3	0	1	1	1	1	1	0	1
			Result:	**0**	**70**	**70**	**37**	**70**	**72**	**21**	**66**

Category	Nr	Requirement	KJ refrences	Weight	Solution 11 (int i app)	Solution 1 (allt i app)	Solution 2 (int dok/web)	Solution 7(PDF)	Solution 10 (1+pop)	Solution 12 (game)	Solution 13 (app)
Main Function	1	Solution shall communicate the annotation task		Must	0						
Sub Function	2	Solution should contain instruction to fullfill the task		5	0	0	0	0	0	0	0
Quality	3	Solution should help enhance quality whilst remaining speed	1.2, 1.3, 1.13,	4	0	0	0	-1	0	0	0
	4	Solution should help customer to formulate what they want annotated/produced	1.4	2	0	0	0	0	0	0	0
	5	Solution should increase probability to cover most scenarios	1.2	3	0	0	0	-1	0	-1	0
	6	Solution should increase probability to cover most edge cases	8.6 13.3	3	0	0	0	-1	0	-1	0
	7	Solution should help minimize annotation time	6.2.3	4	0	1	0	-1	1	1	0
	8	Solution should be effective and efficient to use during production	9.4, 1.14	4	0	1	0	-1	1	0	0
Customer	9	Customer should not need to have insights in the annotation tool	2.1, 2.2	1	0	0	0	0	0	0	0
	10	Customer should not need to be in contact with many different people at	1.5	2	0	0	0	1	0	1	0
Content	11	Should contain relevant examples for the data set being annotated	13.4 13.6	3	0	0	0	-1	0	-1	0
	12	Should contain many examples	12.4, 16.1	3	0	0	0	0	0	-1	0
	13	Should make the user aware of changes	9.5	4	0	0	-1	-1	1	-1	0
	14	Should only use simple english	7.10, 7.15, 13.	4	0	0	0	0	0	0	0
	15	Should express instructions rather than delivery	8.7	3	0	0	0	0	0	0	0
	16	Should have a consistent structure	12.1, 13.8	3	0	0	0	0	0	-1	0
	17	Should have concise information regarding the annotation task and reduce amount of assumptions needed	12.2	3	0	0	0	0	0	-1	0
	18	Lists and tables should be preferred over texts descriptions	16.9	2	0	-1	0	0	-1	0	-1
	19	When referring to something it should be in direct proximity	16.11	3	0	0	0	0	0	0	0
	20	Text, including figure texts, must be readable	16.2	5	0	0	0	0	0	0	0
	21	Length of figure texts should be limited	16.3	3	0	0	0	0	0	0	0
Technicality	22	Should be easy to update/edit		4	0	0	0	-1	0	-1	0
	23	Should be digitally sharable	4.3 5.6 6.1.2	5	0	0	0	0	0	0	0
	24	Should be easy to comprehend regardless of technical background	Brief	4	0	0	0	0	0	0	0
	25	Should be able to quickly find the desired content	12.5, 16.4	3	0	1	0	-1	1	-1	-1
	26	Should provide high resolution pictures	16.2	2	0	0	0	0	0	0	0
	27	Should communicate when a change has been made and the content of that change	13.9	3	0	0	0	0	1	-1	0
Communication	28	Should allow annotators to feel safe to ask any question	7.9	2	0	-1	1	0	0	-1	0
	29	Should allow questions to be raised in an effective and efficient way to and between annotator, BPOQM, AQM, PE and customer	13.5	2	0	0	-1	0	0	-1	0
Other	30	Should have a high level of learnability	10.1 Brief	3	0	0	0	0	0	1	0
	31	Should use prioritization of information and be concise	8.17 8.19	3	0	0	0	-1	0	-1	0
	32	Should minimize the amounts of documents	11.10,	3	0	0	0	-1	0	-1	0
Wishes	33	Should have the ability to include images, GIFs and videos	10.13 12.9 16.	4	0	0	0	-1	0	-1	0
	34	Should encourage to use the solution rather than consulting BPOQM	9.6	3	0	0	0	0	0	0	0
				Result:	**0**	**7**	**-4**	**-40**	**16**	**-37**	**-5**

DEPARTMENT OF INDUSTRIAL AND MATERIAL SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY