DTU Entrepreneurship
DTU Entrepreneurship

Improving crowd safety measurability at Roskilde Festival utilizing AI-enabled video surveillance analysis

Thesis subtitle

Master Thesis

Improving crowd safety measurability at Roskilde Festival utilizing AI-enabled video surveillance analysis
Thesis subtitle

Master Thesis
Date, year

By
Torben Albert-Lindqvist

Cover photo: Vibeke Hempler, 2012
Published by: DTU, DTU Entrepreneurship, Diplomvej, Buildings 371 & 372, 2nd floor, 2800 Kgs. Lyngby Denmark
www.entrepreneurship.dtu.dk/
ISSN: [0000-0000] (electronic version)
ISBN: [000-00-0000-000-0] (electronic version)

ISSN: [0000-0000] (printed version)
ISBN: [000-00-0000-000-0] (printed version)

Approval

This thesis has been prepared over six months at the Section for Indoor Climate, Department of Civil Engineering, at the Technical University of Denmark, DTU, in partial fulfilment for the degree Master of Science in Engineering, MSc Eng.

It is assumed that the reader has a basic knowledge in the areas of statistics.

Torben Albert-Lindqvist - s233587

....

Signature

....

Date

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Acknowledgements

Torben Albert-Lindqvist, MSc Civil Engineering, DTU
Creator of this thesis template.

[Name], [Title], [affiliation]
[text]

[Name], [Title], [affiliation]
[text]

Contents

1 Introduction

1.1 Background and motivation

On June 30th 2000, nine young men lost their lies in a crowd crush during a Pearl Jam concert at Roskilde Festival in Denmark [1]. An uncontrolled surge, pushing the crowd towards the scene, caused immense pressure on the front most concert-goers, thrusting them against the barriers. The high-energy mass of people unknowingly trampled the victims, who succumbed under the pressure of the crowd. This incident is unfortunately not the only one of its kind, as crowd crushes continue to occur at mass gatherings around the world. The 2021 Astroworld Festival in Houston, Texas, resulted in ten fatalities and hundreds of injuries due to a similar crowd surge.

According to NBC News, ”a 56-page event operations plan for the Astroworld festival detailed protocols for dangerous scenarios, including a shooter, bomb or terrorist threats, and severe weather. But it did not include information on what to do in the event of a crowd surge” [2].

1.2 Problem definition

No matter the size of the event, crowd safety is a complex and multifaceted problem. No matter the extent of planning and preparation, the unexpected can still happen. Crowd safety professionals have an immense responsibility, as they are tasked with ensuring the safety of thousands of people. Large crowds are often unpredictable, and improper planning and/or response can lead to disastrous consequences. Fortunately, however, crowd safety professionals have a plethora of tools and knowledge at their disposal to help mitigate these risks and keep crowd dynamics within manageable bounds (see section 2.1).

Despite the available tools and established frameworks, current crowd safety management practices face significant limitations. A recurring theme identified through discussions with industry professionals is a heavy reliance on the experience and intuition of the safety team. Key decisions regarding venue layout, capacity planning, resource allocation (e.g., staffing levels and placement), and risk assessment of concerts often depend heavily on estimations derived from past events and anecdotal knowledge rather than objective, quantitative data specific to the event's context. While experience is invaluable, this reliance introduces subjectivity and potential inconsistencies. Calculations conducted for aspects such as stage layout or entrance/exit dimensions are often based on estimations made by the safety team, albeit these usually are adjusted to cover worst-case scenarios. Even these adjustments, however, are still limited by the imagination and, crucially, the past experiences of the safety professionals. This ultimately suggests that younger or less experienced teams may lack the extensive reference points available to seasoned experts, potentially leading to false assumptions and miscalculations. Consequently, the efficacy of safety planning can appear correlated with the cumulative experience within the team.

This complicated problem, has a seemingly simple and obvious solution: more objective data. At present, experiences and observations often go undocumented, as the safety team already has a plethora of responsibilities and tasks to attend to, not to mention the impossibility of having a full overview of the event at one given moment. Moreover, most music festivals are massively scalable operations, going from a relatively small team of full-time employees and contractors throughout the planning stage, to a massive team of volunteers during the event. In the example of Roskilde Festival, this culminates in a team scaled from over 100 employees to over 30 thousand. This drastic staff scaling certainly applies to the crowd safety department as well, potentially contributing to a lack of continuity and knowledge retention. This implies that immediately following the event, or even after a given day, important observations and learnings from staff may be lost if not systematically documented. These factors effectively hinder a comprehensive, data-driven understanding of crowd behavior.

Furthermore, communicating the rationale behind safety decisions and requirements to other internal departments or external stakeholders can be challenging without clear, objective evidence. Provided that safety precautions are based on subjective assessments, conveying a need for specific resources or precautions can prove difficult. For instance, Roskilde Festival's safety team present a recurring challenge of convincing colleagues in the food and beverage department of concerns regarding the placement of food stalls or bars. Without the aid of clear evidence, Roskilde Festival's safety team occasionally find themselves dedicating valuable time and resources to justify their positions, at times even having to deviate from their core competencies to develop visual material to support their arguments.

These scenarios and limitations highlight opportunities for significant improvement, namely through the development and integration of technology capable of providing objective, measurable insights into crowd dynamics at music festivals. This thesis seeks to explore these opportunities, guided by the following hypotheses:

- Challenges in communicating crowd safety requirements and justifying decisions internally are often due to the subjective nature of current assessments, lacking objective, easily understandable evidence.
- The precision and efficacy of safety planning are constrained by a reliance on experience-based estimations rather than quantitative, historical data on actual crowd dynamics.
- The lack of a persistent, easily accessible digital record of crowd dynamics during an event limits post-event analysis, knowledge retention, and continuous improvement within safety teams.

1.3 Brief history of Fluxense

Together with two classmates, I founded a startup, Fluxense, in January 2024. Our initial plan was to solve the crowd safety challenges of music festivals, as presented in section 1.2, by developing an AI-enabled system for monitoring existing CCTV infrastructure to provide automated analyses of crowd behavior. We gained traction quickly, with several large festivals expressing their interest in our proposed product. Development began almost immediately, and we held our first prototype test at DTU's Commemoration Day, where we provided a live count of the number of people in the concert hall. The test gave great results, as well as valuable learnings, and became the first of many. The following summer was very busy, as we attended three of Denmark's largest music festivals – Copenhell, Roskilde Festival and Smukfest – to further test and develop our product.

After the summer, we stood at a crossroads. Our collaborations with the different festivals had revealed that each had their own unique requirements, and the value of our product was not as clear-cut as we had initially thought. We feared that crowd safety was not a large enough market for scaling our business, nor that a generalized product would be attractive in the industry. We decided to pivot, and began exploring other markets where our technology could be of use. We gradually moved away from our initial focus on crowd safety, and found business intelligence to be a much larger and lucrative market. Instead of monitoring crowds, our new product would track individual customers in retail stores, transportation hubs, amusement parks, and museums. We aimed to provide insights into places/products of interest, dwell times, conversion rate, footfall analysis, etc., to help businesses optimize their operations.

As the autumn progressed, we began securing new collaborations in our target industry, and our value proposition became clearer. One important thing had been lost in the process, however: our motivation. We had started Fluxense with the goal of improving crowd safety at music festivals, as it was a mission we shared a passion for. Our new focus on business intelligence made sense fiscally, but didn't evoke the same feeling of purpose. Fluxense ended up dissolving in the winter of 2024/25, as we couldn't see ourselves in the startup's new reality, and struggled to find a common vision.

1.4 Scope and purpose statement

This thesis continues approximately where Fluxense left off before the pivot. However, rather than following the path laid out by the startup and striving to develop a scalable commercial product, the purpose of this work is to create a tool that directly aids crowd safety managers at Roskilde Festival. The outlined product design and development is exclusively performed for Roskilde Festival, and the result is not intended for commercial use. The focus on Roskilde Festival in particular is partially due to our already close collaboration throughout the entirety of 2024, as well as their expressed interest in continuing our collaboration through this leg of the project. Additionally, it is the largest music festival in Northern Europe, attracting over 130 thousand guests each year [3]. With considerable prestige in the industry, as well as a passionate dedication to improving crowd safety, Roskilde Festival is an ideal partner for this project.

It is important to note that the majority of crowd safety practices presented in this thesis are gathered through discussions with Roskilde Festival's safety team. While occasional references may be made to interactions or insights gained through collaborations with other festivals, these are included solely to illustrate the broader landscape and are not indicative of the project's applicability beyond Roskilde. These findings are also assumed limited to a Danish context, as all discussions with crowd safety professionals were with Danish festivals. Additionally, it was observed during these consultations that some prominent experts occasionally presented viewpoints that could be interpreted as subjective or opinionated. However, as the explicit goal of this project is to provide a functional tool for Roskilde Festival's specific operational environment, a critical evaluation of the objective validity of these statements falls outside the defined scope and is not deemed essential for achieving the project's objectives.

In summary, the purpose statement of this project is as follows:

> *To enhance Roskilde Festival's crowd safety management by developing an intuitive, data-driven platform that provides actionable insights into crowd dynamics, thereby improving planning, internal communication, and documentation for the safety team.*

1.5 Objectives

Formalizing the hypotheses presented in section 1.2, the objectives of this project are as follows:

1. **Improve internal communication**: Offer clear, visual, objective evidence to help the safety team communicate requirements and justify decisions to other departments.
2. **Enhance safety planning**: Provide quantitative, historical data on crowd dynamics to enable more accurate planning of layouts, capacities, resource allocation, and facility placement.
3. **Create reliable documentation**: Generate a persistent digital record of crowd dynamics for post-event analysis, debriefing, and knowledge retention.

1.6 Thesis structure

In their book, *Design Science*, Hubka and Eder characterize the design process as intuitive, iterative, innovative, unpredictable and reflective [4]. While these aspects are inherent to design, tackling complex engineering challenges requires more than intuition and creativity alone. To manage the process effectively, ensure thoroughness, and facilitate clear understanding and traceability, a structured approach is beneficial [5]. Therefore, the process needs to be organized, drawing upon established product design methodologies and frameworks to guide this project. Many such frameworks exist, offering different levels of detail and focus. This section will explore the most relevant frameworks, and propose a design and development methodology for this project.

1.6.1 Comparison of frameworks

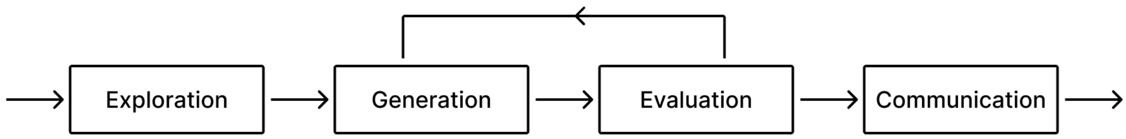


Figure 1.1: Cross' four-stage model of the design process

Cross [6] proposes likely the most simplistic, yet well-known framework: a four-stage model comprised of *exploration*, *generation*, *evaluation* and *communication* (Figure 1.1). Cross describes this type of model as descriptive, as it merely attempts to model the conventional, heuristic design process. More detailed models of this type exist, such as French's [7] "anatomy of design," (Figure **??**) detailing four stages, most distinctly underlining the problem analysis and definition, as conducted in section 1.2. According to Cross, these descriptive models differ from prescriptive ones, which offer a more systematic procedure, as well an emphasis on analyzing and understanding the design problem before generating solution concepts. Perhaps the most well-known of these is offered by Pahl et. al [8], and is based on the following design stages: *clarification of the task*, *conceptual design*, *embodiment design*, and *detail design*. Combining many of elements present in the aforementioned models, Ulrich and Eppinger present a rather comprehensive framework. Their process is based on the following stages: *concept development*, *system-level design*, *detail design*, and *testing and refinement* [9]. Concept development involves understanding the needs of the target user, and clearly defining the problem, as well as developing prototypes for further testing. The system-level design stage focuses on the architecture of the product, and defining the subsystems and components. Subsequently, the detail design takes the components from the previous stage, and defines the requirements and specifications required for manufacturing. This followed with testing and refinement, and ultimately the production of the product.

These frameworks provide varying degrees of structure and granularity to the design/development process, but all share the commonality of being highly engineering-focused. In engineering a physical product, a rigid, structured process is often necessary as each iteration must be designed, manufactured and tested. This is costly, both in effort and material costs. Therefore, the design and development process are divided and sequential. Software, on the other hand, is much more flexible, with iterations being a magnitude faster and cheaper to develop. This demonstrates a need for adapting the design/development process to the context of the product being developed. Conveniently, Ulrich and Eppinger present a multitude of adaptations to their framework, including what they refer to as ”Quick-Build Products” and ”Digital Products.” Here the *detail design* and *testing and refinement* stages are omitted, and replaced with a cyclical design-build-test process. Whereas the linear, rigid processes described previously are labelled as ”waterfall methods”, this iterative process is most often referred to as *agile development*.

Agile development has many benefits in contrast to the waterfall approach, especially in the context of software development. As mentioned previously, the waterfall approach is ideal for engineering projects where prototyping is costly. When the cost of prototyping is negligible, however, agile methodology grants the flexibility to iterate quickly, and adapt to evolving requirements. Design and development are sequential in a waterfall model; here they are heavily intertwined. A strong example of this, as well as being the most popular implementation of agile development, is *Scrum*. Scrum defines the following stages: *sprint planning*, *daily stand-up*, *sprint review*, and *sprint retrospective*, with a sprint typically lasting 2-4 weeks [10]. This framework is ideal for large teams, as it ensures that all team members are aligned, and are able to coordinate their efforts efficiently. The daily stand-up is a particularly useful tool for larger teams, preventing overlapping work, or potential blockers from being overlooked. In smaller teams, however, this structure can be cumbersome, and potentially even counterproductive. Especially when considering this project, exploring a singular use-case as a solitary developer, the full-scale implementation of Scrum is evidently not necessary. Instead, a more adaptable and lightweight framework is employed.

In his book, *The Lean Startup*, Eric Ries describes a simple, yet effective agile framework, which he refers to as the *Build-Measure-Learn* loop [11]. This framework is designed for rapid prototyping and iteration, splitting each stage into *build*, *measure* and *learn*. *Build* involves developing a minimal product or feature, which is then tested with the target user(s) in the *measure* stage. The results of this test are then studied in the *learn* stage, where the product/feature is adapted based on these results. This process is then repeated, until the requirements are met. This framework is ideal for this project, as it is quite exploratory in nature, while still aiming to fulfill predetermined requirements.

1.6.2 Design and development methodology

Building upon the exploration of frameworks conducted above, the following design methodology is proposed for this project. The design stage follows an engineering approach, guided by Ulrich and Eppinger’s *concept development* and *system-level design* stages. Subsequently, the development stage is based on a more agile, entrepreneurial approach, as described by Ries. Standing in for the *detail design* and *testing and refinement* stages, Ries’s *Build-Measure-Learn* architecture is employed in order to facilitate rapid prototyping and iteration. Each iteration includes a *build* stage, where a new set of features are developed, followed by a *measure* stage, where the new features are presented to the target user, and feedback is gathered. This is followed by a *learn* stage, where the feedback is analyzed, and the next set of features are defined. This

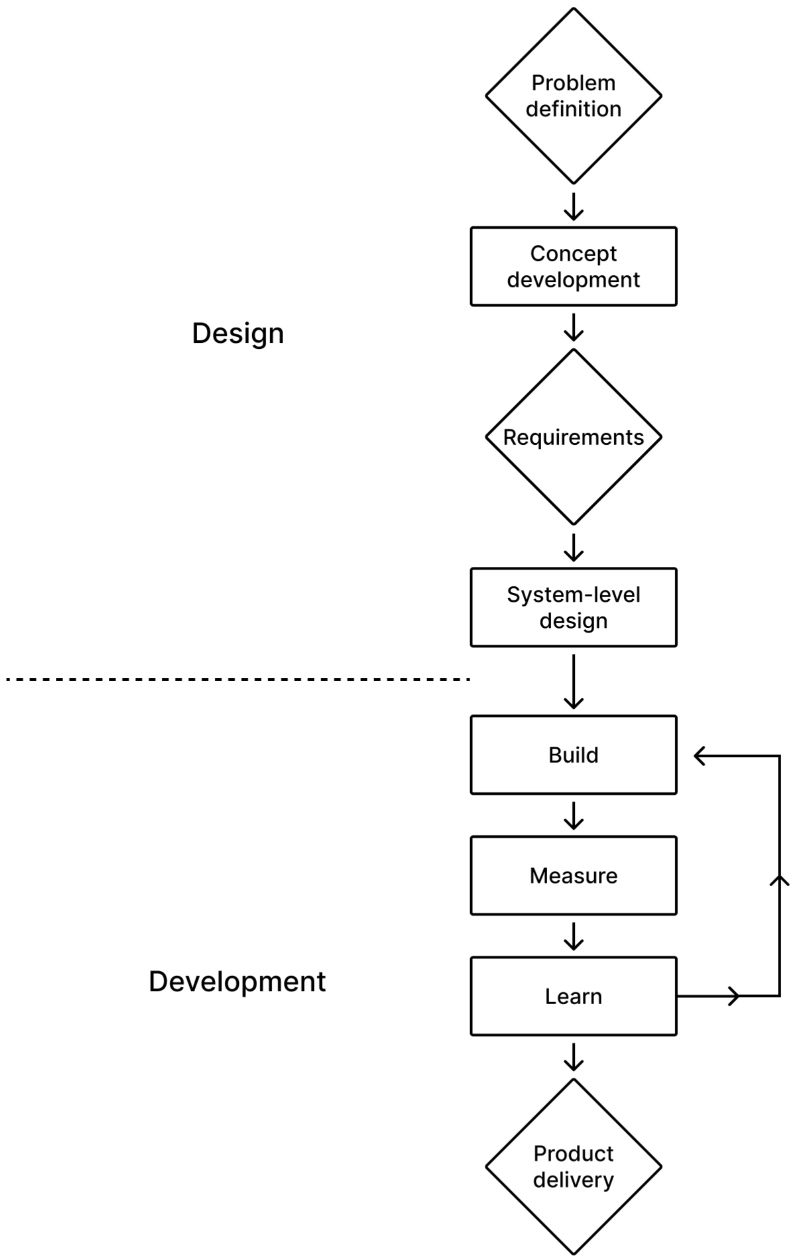


Figure 1.2: The proposed design and development framework for this project. Each diamond indicates a crucial stage or milestone, while each box represents a stage in the design and development process. The design stages follow Ulrich and Eppinger's framework, while the development stages are based on Eric Ries' *Build-Measure-Learn* framework.

process is repeated until the requirements are met (Section 2.3). Figure 1.2 illustrates the proposed design and development framework.

Chapter 2, **Concept Development**, ultimately selects and proposes a conceptual solution to the problem outlined in section 1.2. As defined by Ulrich and Eppinger, a concept is "a description of the form, function, and features of a product and is usually accompanied by a set of specifications, an analysis of competitive products, and an economic justification of the project." [9] This selection is initially preceded by a thorough inspection into the intricacies of crowd safety management, as well as a review of potential solutions and existing products. A novel solution is thereafter presented, outlining requirement specifications, as well as its legal and financial feasibility.

Chapter 3 focuses on the **System-level Design**, including the definition of the product architecture, and a decomposition of the product into subsystems and components. The full workflow from data collection to the resulting user interface is presented, followed by a detailed description of each sub-system, including the data collection, computer vision model, spatial mapping, metric extraction, and the user interface/frontend.

Finally, Chapter 4 presents the results of the development stage, as dictated by the **Build-Measure-Learn** framework. The chapter begins with a showcase of the resulting application and brief summary of feature selection. Following this, a technical performance evaluation is conducted, as well as an evaluation of the product's business value through a workshop with Roskilde Festival. Finally, the chapter concludes with a discussion of the results, as well as a reflection on the product's value and fulfillment of its requirements.

2 Concept Development

2.1 Understanding crowd safety management

In order to develop a solution that supports crowd safety professionals, it is imperative to understand how they operate, and what tools they currently have at their disposal. Together with my co-founders at Fluxense, we conducted interviews with many crowd safety professionals from various different organizations, including Event Safety (Smukfest), smash! bang! pow! (Syd for Solen), Roskilde Festival, and Live Nation (Copenhell, Heartland). Throughout this period, it became clearer that crowd safety management is very complex, and is almost as much a philosophy as it is a science. Music festivals and events vary greatly in size, participant demographics, venues, and budget. Equally varied are the crowd safety professionals themselves, who appeared to have varying levels of experience, as well as distinct approaches to their work.

Most interestingly, the greatest discrepancy was seemingly between a focus on incident-prevention and incident-response, or ”crowd safety vs. security”, as according to Roskilde Festival’s Director of Safety, Morten Therkildsen (Appendix A.2.2). A security-focused approach often involves less planning, as well as hiring third-party professionals to handle safety during the event. Safety-focused teams, on the other hand, spend most of the year leading up to their events meticulously planning initiatives to ensure the well-being and enjoyment of their guests. The distinction between these two protocols was apparent throughout Fluxense’s collaborations with both Copenhell and Roskilde Festival. At their 2024 events, Live Nation had two full-time employees responsible for crowd safety at Copenhell, whereas Roskilde Festival had a team of 10+ full-time employees. Additionally, our collaboration with Copenhell was focused on real-time analysis, while Roskilde Festival is much more interested in post-event analysis.

This difference in approach is reflective of Denmark’s regulatory landscape regarding crowd safety. While official documentation exists, such as the ”Vejledning om sikkerhed ved udendørs musikarrangementer o.lign.” published by the Ministry of Justice and Ministry of Culture, these serve as guidelines rather than enforceable legislation specifically covering the planning phase of crowd safety management. The document itself emphasizes its role as a tool and catalog of ideas for organizers, who ultimately retain the responsibility for risk-assesment and implementation of appropriate measures based on their specific event [12]. Furthermore, the Danish Police have published supplementary guidance, describing the requirements for receiving a permit. While the police require organizers of large events to submit a safety plan as a condition for obtaining the necessary event permit, this is solely focused on incident-response [13]. Thus, although numerous legal requirements touch upon event safety, there isn’t distinct legislation dictating the specific process and minimum standards for proactive crowd safety planning. The legislation highlighted in these guidelines focuses on structural safety, emergency response, and police involvement, leaving the interpretation and extent of proactive safety planning largely to the individual organizers. This contributes to the observed discrepancies in how crowd safety management is implemented across different Danish festivals, as they interpret the available guidelines and integrate them into their operational workflows individually.

As the scope of this thesis is to develop a solution for Roskilde Festival explicitly, their distinct interpretation of crowd safety will be the exclusive focus in requirement gathering.

The following sections will outline the current frameworks and workflows used by Roskilde Festival, as well as the key metrics they calculate and monitor.

2.1.1 Existing frameworks and workflows

Roskilde Festival's approach to crowd safety management is built upon internationally recognized frameworks, prioritizing proactive planning and analysis. Roskilde Festival's Director of Safety, Morten Therkildsen, considers the United Kingdom to be on the forefront of crowd safety management, and the festival follows several UK-based frameworks to inform their practices. These include *The Purple Guide* [14], as well as the methodologies of Professor G. Keith Still, whose work highlighted in *Applied Crowd Science* has been adopted for mandatory training for public event commanders by the UK College of Policing since 2018 [15]. Roskilde Festival also references the *Event Safety Guide* from the US-based Event Safety Alliance (ESA) [16]. Roskilde's practical implementation of these frameworks is apparent in many of their workflows, several of which are outlined below.

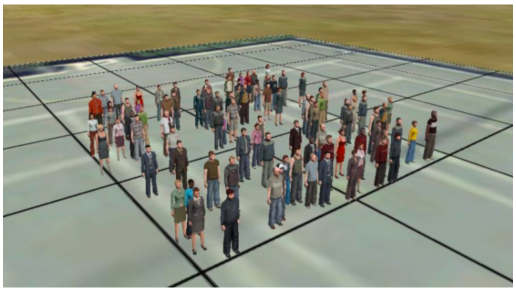
An integral workflow involves conducting artist- and stage-specific risk assessments for every concert. The reasoning is to anticipate potential hazards based on the unique characteristics of each artist and their expected audience. This involves a detailed "band analysis" (Appendix A.2.2) considering factors like genre, typical crowd behavior, demographics, together with Still's RAMP analysis (evaluating routes, area, and movement specific to the stage, as well as the people/audience specific to the artist) [17]. Subsequently, concerts are categorized using a system reflecting risk based on the RAMP analysis (red/yellow/green), and expected attendance relative to stage capacity (A: <25%, B: 25%-75%, C: 75%-100%, C+: 100%+). This assessment enables the safety team to tailor resources – such as staffing levels, barrier placement, and egress/flow management – to the predicted risk profile of each concert.

Understanding and managing how attendees move is addressed through flow analysis. Roskilde Festival's safety team analyze ingress/egress patterns and rates, particularly at entrances/exits, aggregating data across various time intervals (1, 15, 60 minutes). The objective is to prevent bottlenecks and ensure that the physical infrastructure can safely accommodate peak movements. This quantitative understanding of flow is crucial for accurately calculating required pathway widths, optimizing crowd control measures like queue management, and deploying staff to guide attendees when necessary. Additionally, crowd density analysis, using metrics like Levels of Service (LoS) inspired by John J. Fruin [18], is employed to monitor crowd concentration (see Roskilde Festival's implementation in Figure 2.1). This is especially conducted in areas with expected high density (e.g., front-of-stage) or potentially congested zones (e.g., queues around vendors). Its purpose is to prevent the high densities that pose a direct risk of crowd crushes, and to evaluate the effectiveness of the site layout in distributing the crowd effectively.

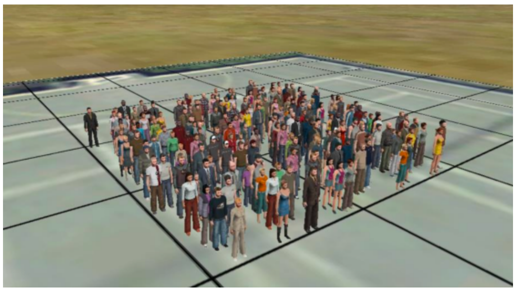
2.1.2 Key metrics

Given the workflows outlined above, we can identify the following key metrics that Roskilde Festival relies on to assess crowd safety:

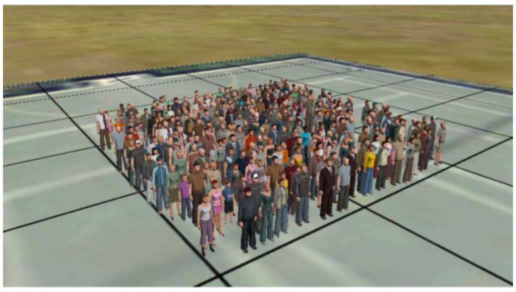
- **Ingress/Egress Counts**: The number of people entering (ingress) or leaving (egress) a monitored area.
- **Flow Rates**: The rate at which people enter or exit an area over specific time periods (e.g., per minute, per hour), including net change and peak rates.



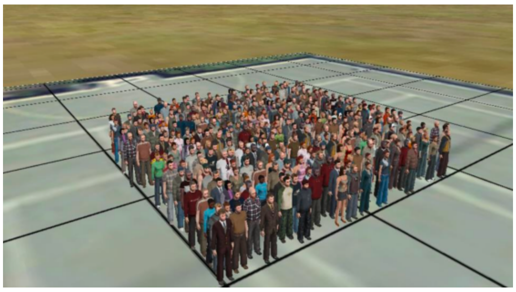
A (0-1 person/m^2). Head, shoulders, chest, and feet are visible.



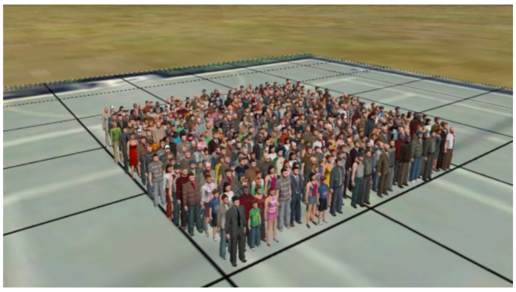
B (1-2 people/m^2). Head, shoulders, chest, and feet are visible.



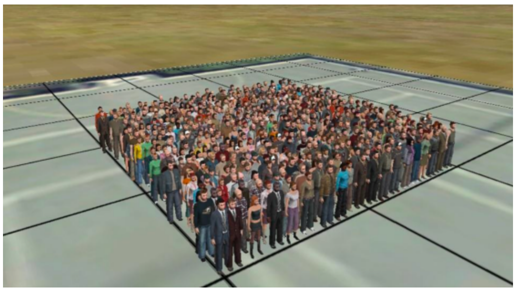
C (2-3 people/m^2). Head, shoulders, and chest are visible.



D (3-4 people/m^2). Head, shoulders, and chest are visible.



E (4-5 people/m^2). Head and shoulders are visible.



F (5+ people/m^2). Only heads are visible.

Figure 2.1: Roskilde Festival's internal implementation of Levels of Service (LoS), using a scale of A-F. The images serve as visual examples of the different LoS levels, used to estimate crowd density on-site or through video footage. Roskilde relies on visual cues, namely the visibility of heads, shoulders, and feet, to assess crowd density.

- **Cumulative Counts**: A running total showing the net number of individuals within the monitored area over time.
- **Crowd Density**: The concentration of people within an area, typically expressed as people per square meter (people/m^2).
- **Movement Patterns**: The general paths, directions, origins, and destinations of people moving within the monitored space.

2.1.3 Revisiting the problem definition

As outlined in the previous sections, Roskilde Festival currently utilizes established frameworks and workflows for proactive crowd safety management. Their approach demonstrates a commitment to data-driven planning and analysis. However, a significant challenge lies not in the *analysis* of data, but in the *acquisition* of reliable and objective data. While the festival's safety team already has the tools necessary to interpret and act upon key metrics, the current means of gathering this information often rely on manual, time-consuming, or subjective methods.

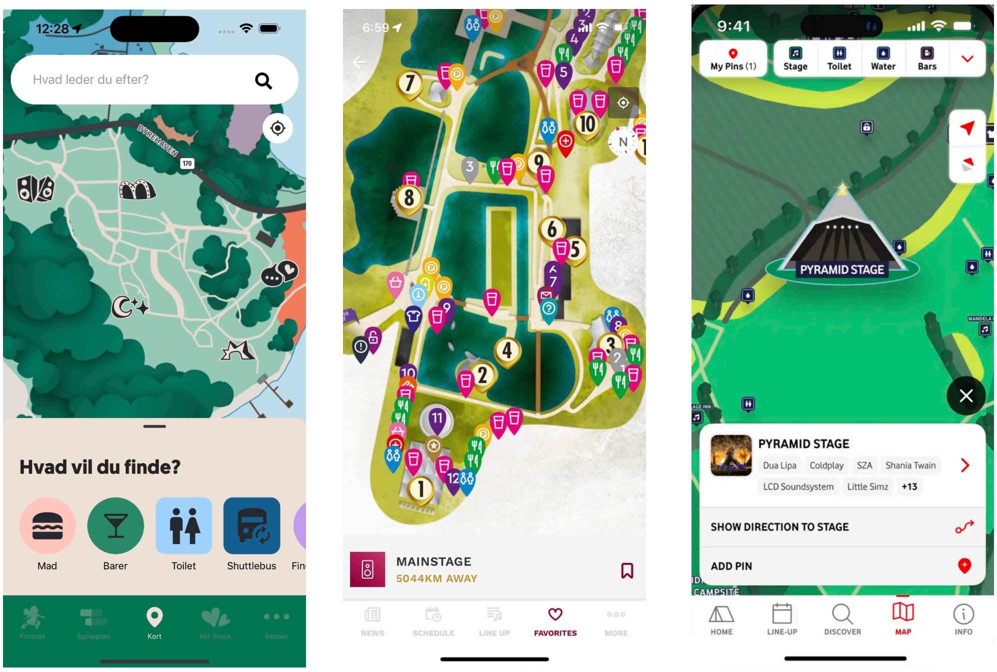
A prime example is the assessment of ingress/egress flow rates. As described by Mads Therkildsen, Safety Lead at Roskilde Festival, a common practice involves manually counting individuals passing through an entrance for a short period, such as one or two minutes, and then extrapolating this count to estimate hourly flow (Appendix A.2.6). While practical, this method provides only partial data and is likely inaccurate, especially during rapidly changing conditions. Similarly, crowd density is currently gauged primarily through subjective visual assessment using an internal Levels of Service (LoS) scale, as illustrated in Figure 2.1. The current process relies on observers estimating density based on visual cues like the visibility of heads, shoulders, and feet. Consequently, the true density value across different areas and times remains largely unknown.

These examples highlight the core limitation: the absence of an efficient system to autonomously and objectively gather quantitative data on key crowd metrics. As Roskilde Festival already employs a data-driven approach to crowd safety management, such a system would integrate directly into their existing workflows. The following sections explore potential technical solutions to solve this challenge, eventually proposing a product.

2.2 Comparing technical solutions

2.2.1 Global Positioning System (GPS)

Using GPS to track the location of festival-goers is a common practice, and likely the easiest to implement technology in this comparison. This is typically achieved by providing guests with a mobile app that uses their smartphone's GPS to track their location. Evidently, this requires the guests to opt in to location tracking, as well as there being a sufficient reason for doing so. In almost all cases, this is attempted by including a map of the festival in the app, as visible in Figure 2.2. This feature, however, still functions without location tracking, and therefore doesn't guarantee users will grant data access. Even before this obstacle is met, there is the question of whether festival-goers will actually use the app. A study on Roskilde Festival 2015 by the Copenhagen Business School found that of the 60 thousand people who installed the festival application, 44 thousand opted-in to allowing anonymous tracking; yielding 38.678 unique users who were present inside the festival area [19]. This equates to slightly under 30% of the total 130 thousand attendees. In a crowd safety context, this is a significant limitation, as the location data gathered is not representative of entire crowds.



(a) Smukfest app [20] (b) Tomorrowland app [21] (c) Glastonbury app [22]

Figure 2.2: Examples of map features in music festival mobile apps (Smukfest, Tomorrowland, Glastonbury), often used to encourage attendees to enable GPS location tracking.

Beyond low adoption and potential privacy concerns, the technical limitations of GPS also hinder its utility for detailed crowd analysis. According to GPS.gov, GPS-enabled smartphones are typically accurate only to within a 4.9-meter radius under open sky; however, their accuracy worsens near buildings, bridges, and trees [23]. While a 4.9-meter radius might seem acceptable for general location awareness on a festival map, this level of uncertainty significantly hinders the calculation of precise crowd dynamics metrics. Furthermore, the degradation of accuracy near structures is particularly problematic in festival environments, which often feature large stages, tents, and temporary structures – precisely where accurate monitoring is most needed. The effectiveness of GPS tracking is also contingent on factors outside the organizers' control, such as users keeping their phones charged and maintaining a stable mobile data connection.

Compared to infrastructure-based monitoring systems (like cameras or dedicated sensors), GPS relies heavily on user cooperation and device functionality, making it less suitable for generating the consistent, high-resolution data needed for proactive crowd safety management and detailed post-event analysis. Therefore, while mobile app GPS data can offer some high-level insights into general attendee distribution, its inherent limitations in accuracy make it insufficient as a primary tool for gathering crowd dynamics measurements.

2.2.2 Bluetooth beacons

Another potential solution for gathering crowd dynamics data involves utilizing Bluetooth Low Energy (BLE), a wireless communication technology designed for low power consumption. This technology is typically utilized in Bluetooth beacons – devices that periodically transmit signals that can be detected by nearby receivers, such as smartphones or dedicated hardware. These signals can be used to determine the proximity of beacons to the receivers, as well as location estimation through triangulation techniques, requiring three or more receivers. This capability is increasingly utilized for proximity marketing, indoor navigation, and positional tracking, typically in retail environments [24]. It has also been effectively used for guest tracking in environments like museums, as demonstrated in a study at the Galleria Borghese, where visitors were given portable BLE beacons tracked by fixed receivers [25].

Finding examples of BLE beacons utilized at large outdoor events is more challenging, as the technology has its own set of drawbacks. Similar to GPS-based mobile apps, the effectiveness of Bluetooth beacon systems is significantly contingent on user cooperation. Beacons can be implemented either as dedicated physical devices or embedded within smartphone applications. The latter is simpler and cheaper, however motivating users to install an app is challenging, as discussed in Section 2.2.1. The former approach has challenges of its own, as requiring event participants to carry a dedicated device demands a value proposition of its own, not to mention the overhead costs involved in developing, deploying, and maintaining the system. Given that beacon signals have an effective range of 30 meters without obstructions, a considerable number of receivers are necessary to ensure adequate coverage [24].

Ultimately, the most critical factor for crowd dynamics analysis is the technology's accuracy. A study conducted at a large exhibition in Japan evaluated the accuracy of trajectory estimation using Bluetooth beacons. They found that in 68% of instances, their positional estimates were accurate within a radius of 18 meters [26]. While still useful for certain applications like indoor tracking or general zone classification, this level of accuracy is significantly less precise than typically achievable with GPS (Section 2.2.1). This limitation, combined with the dependency on user adoption and implementation costs, makes Bluetooth beacons a less favorable option for gathering crowd dynamics data at large outdoor events.

2.2.3 Camera solutions

Camera-based solutions leverage existing CCTV systems in conjunction with computer vision algorithms to analyze crowd dynamics. Utilizing video footage from these camera systems, novel advancements in machine learning and computer vision enable automated extraction of valuable metrics.

A primary advantage of this approach compared to GPS and Bluetooth beacon systems is its independence from attendee cooperation. Unlike solutions requiring festival-goers to install an app, enable location services, or carry a dedicated device, camera-based analysis operates passively. It gathers data from anyone within the camera's field of view, potentially offering a more comprehensive and unbiased view of crowd behavior across monitored areas. This circumvents the significant limitations associated with low adoption rates and opt-in requirements inherent in the other methods.

While various companies offer camera-based analytics for large events, not all are directly aligned with the specific challenges intrinsic to crowd safety at music festivals like Roskilde Festival. For instance, Remark AI provides solutions for smart stadiums and events, focusing broadly on operational efficiency, queue management, and general security features [27]. Similarly, Lumeo offers a platform for event analytics with functionalities like occupancy counting, wait time analysis, and object-left-behind detection [28]. However, neither Remark AI nor Lumeo appears to prioritize in-depth crowd safety metrics in their central offering, nor do they showcase any track record of collaboration with major festival environments where complex crowd dynamics are a primary concern. Their broader focus on general operational and security aspects, rather than specialized crowd safety analysis for large, dynamic outdoor events, makes them less directly comparable to the aims of this project.

In contrast, Dynamic Crowd Measurement (DCM) present a solution that demonstrates a considerable overlap with the objectives of this thesis. DCM offers real-time crowd analytics, including density, flow speed, headcount, sentiment and demographic characteristics, aimed at various sectors like events, smart cities, and transport hubs [29]. Despite their seeming relevance to the project, upon closer inspection, it becomes evident that their solution falls short in various aspects. Firstly, DCM's focus on real-time analytics, while valuable in certain contexts, does not align with Roskilde Festival's primary interest in post-event analysis. Roskilde's safety team has expressed a strong preference for analyzing historical data to inform future planning and decision-making, rather than relying on real-time action (Appendix A.2.1). Secondly, DCM's solution doesn't appear to track an area's capacity and flow by tracking ingress and egress counts, and instead only provides a snapshot of the headcount detected by the camera at a given moment. This is a significant limitation, as the importance of this data is outlined in Section 2.1.1. Thirdly, DCM appears to be shifting focus in their technological development towards detecting crowd demographics, characteristics, and sentiment, as evidenced by their recent patent, which details systems for identifying such attributes, with no mention of crucial crowd safety metrics from Section 2.1.2 [30]. Finally, and likely most crucially, DCM's applicability in a European context is questionable, as they are based in Australia and do not explicitly state their compliance with European data protection regulations.

The landscape of camera-based solutions clearly demonstrates room for improvement. While existing solutions like DCM offer some relevant features, they do not fully align with the specific needs of Roskilde Festival's safety team. The following section introduces the proposed solution, which aims to address these gaps and provide a comprehensive, compliant, and user-friendly platform for crowd safety analysis.

2.3 Proposed solution

The proposed solution is an AI-enabled video surveillance analysis platform designed to improve crowd safety measurability and management specifically for Roskilde Festival. The product is designed to operate through a sequence of integrated systems. Initially, data collection is performed, in this case using strategically positioned cameras, however existing infrastructure can also be utilized. This is supported by Roskilde Festival's statement that CCTV cameras are already placed at points of interest from a crowd safety perspective, such as entrances and exits. The captured footage is subsequently processed by a computer vision model, which achieves head detection and individual tracking. Spatial mapping translates these detections from the camera's 2D perspective into real-world geographic coordinates. Given this positional data, key metrics can be extracted. Finally, these insights made accessible to the Roskilde Festival safety team via a web interface, designed for intuitive data exploration and visualization.

Formalizing the requirements for the proposed product is a crucial step in the development process. In each stage of the development cycle, and more crucially, in the final product, these requirements will serve as a benchmark for evaluating the product's progress. The requirement specifications, derived from the needs identified in collaboration with the Roskilde Festival safety team, are as follows:

1. **Core metric extraction**: The product must automatically process inputted video data and accurately extract key crowd metrics. See included metrics in Section 2.1.2.

2. **Intuitive user interface**: The product must provide a user-friendly web interface that presents the extracted data through clear and interactive visualizations. Users (Roskilde Festival safety team) must be able to:
   - Easily navigate the application.
   - Easily access and interpret the data.
   - Filter data by date, time, and specific monitored locations/entrances.
3. **Compliant data handling**: The system must handle input video data and derived analytical data securely, ensuring compliance with relevant data protection regulations, such as the General Data Protection Regulation (Section 2.4). Specifically:
   - All processing of identifiable data (video footage) must lead to anonymized results (e.g., counts, density maps, aggregated flow data) for storage and presentation.
   - Access to the platform must be secured via user authentication.

2.4 Legal Feasibility

When dealing with CCTV footage, there are a number of legal considerations to take into account. This section outlines the primary legal considerations for the project, focusing on Danish CCTV law, the General Data Protection Regulation (GDPR), and the EU AI Act. As this section focuses on the legal feasibility of the product itself, the theoretical administrator will be referred to as Fluxense, the startup mentioned in Section **??**. This name was also used in the Non-Disclosure Agreement (NDA) signed with Roskilde Festival, which underscores the importance of data confidentiality and adherence to applicable laws.

The Danish Act on CCTV Surveillance (tv-overvågningsloven) regulates the use of surveillance cameras in Denmark, which is evidently relevant to this project. The Act generally prohibits private individuals and organizations from conducting surveillance in public areas of ”general traffic,” such as streets and squares [31]. In the context of Roskilde Festival, however, this legislation does not apply, as it is not considered a public area. In these cases, the Act refers to data protection legislation, as outlined by The Danish Data Protection Agency (DDPA) and the European Union’s GDPR. GDPR establishes a comprehensive framework for protecting individuals’ privacy and personal data, including CCTV footage. Both the DDPA and GDPR present two critical roles: the data controller and the data processor. The data controller is the entity that determines the purposes and means of processing personal data, while the data processor is responsible for processing data on behalf of the controller [32]. Both roles are formally defined in a data processing agreement (DPA) between the two parties. In this case, Roskilde Festival acts as the data controller, while Fluxense is the data processor. As data processor, Fluxense is responsible for ensuring that all data processing activities comply with GDPR regulations. This includes ensuring secure processing, containment of sensitive data, as well as prompt deletion of data at the request of the data controller (GDPR Article 28 [33]).

The EU AI Act establishes a regulatory framework for artificial intelligence systems based on a risk-based approach, which is relevant to the AI-enabled video analytics central to this project. The legislation entered into force on August 1, 2024, and will be applicable in stages over the coming months and years [34]. The Act classifies AI systems into four risk categories: unacceptable risk, high risk, limited risk, and minimal risk. As documentation pertaining to the explicit categorization criteria will not be available until August 2026, it is difficult to speculate on the risk classification of the proposed system. It is reasonable to assume that it would not be categorized as unacceptable risk as the AI is not performing any biometric identification, categorization of people or facial recognition. Given the worst case scenario that it is classified as high risk, the following requirements defined by the AI Act would apply: establishing a risk management system, ensuring data governance and management, maintaining technical documentation, transparency with users, ensuring human oversight, and ensuring robustness, accuracy, and security [35]. The proposed system does not currently meet these requirements, however it is a reasonable feat to comply with them in the future.

2.4.1 Financial Feasibility

Although the Section 1.4 makes a point to outline that the product is not intended for commercial use, it is still important to consider its financial viability. The final product presented in this project is delivered to Roskilde Festival free of charge, yet given the product proves a success and Roskilde Festival is interested in a continued collaboration, this theoretical aspect of business feasibility is relevant. Given the product has some costs involved, and assuming that both parties simply desire to ”break even,” the price of the proposed product should not exceed the costs saved by Roskilde Festival. In other words, it must be a business positive solution.

No information was gathered on the Roskilde Festival's budget, nor their costs/time investment associated with crowd safety management. Crucially however, Morten Therkildsen, Roskilde Festival's Director of Safety, provided a two price bounds for a product like the one I propose. He stated that 20,000 DKK would be a fair price, and 100,000 DKK wouldn't even get us in a room together. This serves as a useful benchmark for the product's financial feasibility by providing an estimated price range: +/- 20,000 DKK.

Determining the product's approximate price is simply a matter of estimating the costs associated with developing and maintaining the product, as goal is not profitability for either party. For the purposes of developing and testing the product, free services are sufficient, however an increase in data volume or users will likely require paid services. Table 2.1 outlines the estimated costs associated with the yearly maintaining the product, as well as running analyses on a new round of footage. The estimates are gathered from Microsoft Azure's pricing calculator [36]. The costs are based on worst-case assumptions, such as inference taking twice as long as expected and the application being used 80 hours per month (4 hours per day, 5 days a week), on average over a yearly period. The resulting estimated yearly cost is 7,176.92 DKK, which is well within the bounds set by Morten Therkildsen.

To take this calculation one step further, we assume the average salary of a Safety Lead is 53,650 DKK per month, as reported by Jobindex [**jobindex**]. This equates to an hourly wage of 357.65 DKK, meaning that the product must save Roskilde Festival at least 20 hours of work per year to be considered financially feasible ($7176.92 \div 357.65$).

Service	Yearly Cost (DKK)	Purpose
App service	1,358.67	Hosting the web application
SQL database	4,931.96	Hosting the database for data storage
Virtual machine - NVIDIA A100 - 96 cores - 880 GB RAM	886.29	One-time cost of training models and running inference on new footage
Total	**7,176.92**	

Table 2.1: Yearly cost estimation of the proposed product. The costs are based on Microsoft Azure's pricing calculator [36]. The yearly cost of the virtual machine is a one-time cost, as it is only used for training the models and running inference on new footage.

3 System-level Design

This chapter provides a comprehensive overview of the system-level design, as defined by Ulrich and Eppinger [9]. After presenting the product architecture, the chapter delves into the individual sub-systems that comprise the system. Each sub-system is described in detail, including its purpose, functionality, and how it integrates with the overall system.

3.1 Product architecture

3.2 Sub-systems

The system is composed of five main sub-systems, each serving a distinct purpose in the overall functionality of the product. These sub-systems are: *Data collection*, *Computer vision model*, *Spatial mapping and GIS*, and *Metric extraction*, and finally the *Interface/frontend*. Each sub-system is sequentially dependent, as illustrated in Figure 3.1.

3.2.1 Data collection

Data collection constitutes the initial, vital stage of the system, providing the raw data required for all subsequent processing and analysis. This process entailed the on-site deployment of designated cameras, strategically positioned to capture the targeted crowd dynamics.

The selected cameras were Reolink RLC-520A, which are PoE-enabled (Power over Ethernet) and capable of recording 5MP (2560x1920 pixels) video at 30 frames per second. A separate PoE switch (Ubiquiti PoE++ Adapter), connected to a standard power outlet, was required to power the cameras. This setup allowed connection to the cameras via an Ethernet cable linked to a laptop, enabling camera configuration and live feed monitoring. Utilizing the cameras' integrated software, recording windows could be predefined such that footage would automatically be archived on the internal SD card. This setup was designed to ensure that the cameras could operate independently without requiring a constant connection to a computer. Mounting the cameras was achieved with 3D-printed brackets, designed to securely attach the cameras to existing infrastructure, such as fences or poles. Where existing structures were unavailable, aluminum poles were utilized to achieve the necessary height for capturing the entire designated area. Four cameras were deployed, denoted as *CAM1*, *CAM2*, *CAM3*, and *CAM4*.

As agreed upon with Roskilde Festival's safety team, the cameras were installed around the Eos stage during the first three days of the festival, and subsequently moved to the Arena stage for the remainder of the festival.

Determining camera placement for the Eos stage was relatively trivial. During the festival's "First Days," (June 30th to July 2nd) the remainder of the festival site, excluding the Gaia stage, was closed off to guests, leaving only two pathways for entering and exiting the stage area. At each of these two pathways, two cameras were installed, oriented to face one another. *CAM2* and *CAM3* were positioned at Eos' southern entrance, Entrance 10, while *CAM1* and *CAM4* were placed at the eastern entrance, towards the Gaia stage (Figure 3.2). This dual-camera configuration served two purposes: ensuring complete monitoring of the pathway's width, as well as providing a redundant dataset for each location, effectively mitigating the risk of equipment malfunction during the initial deployment.

Camera placement for the Arena stage presented greater complexity due to its significantly larger scale and increased number of entrance and exit points. Arena is located

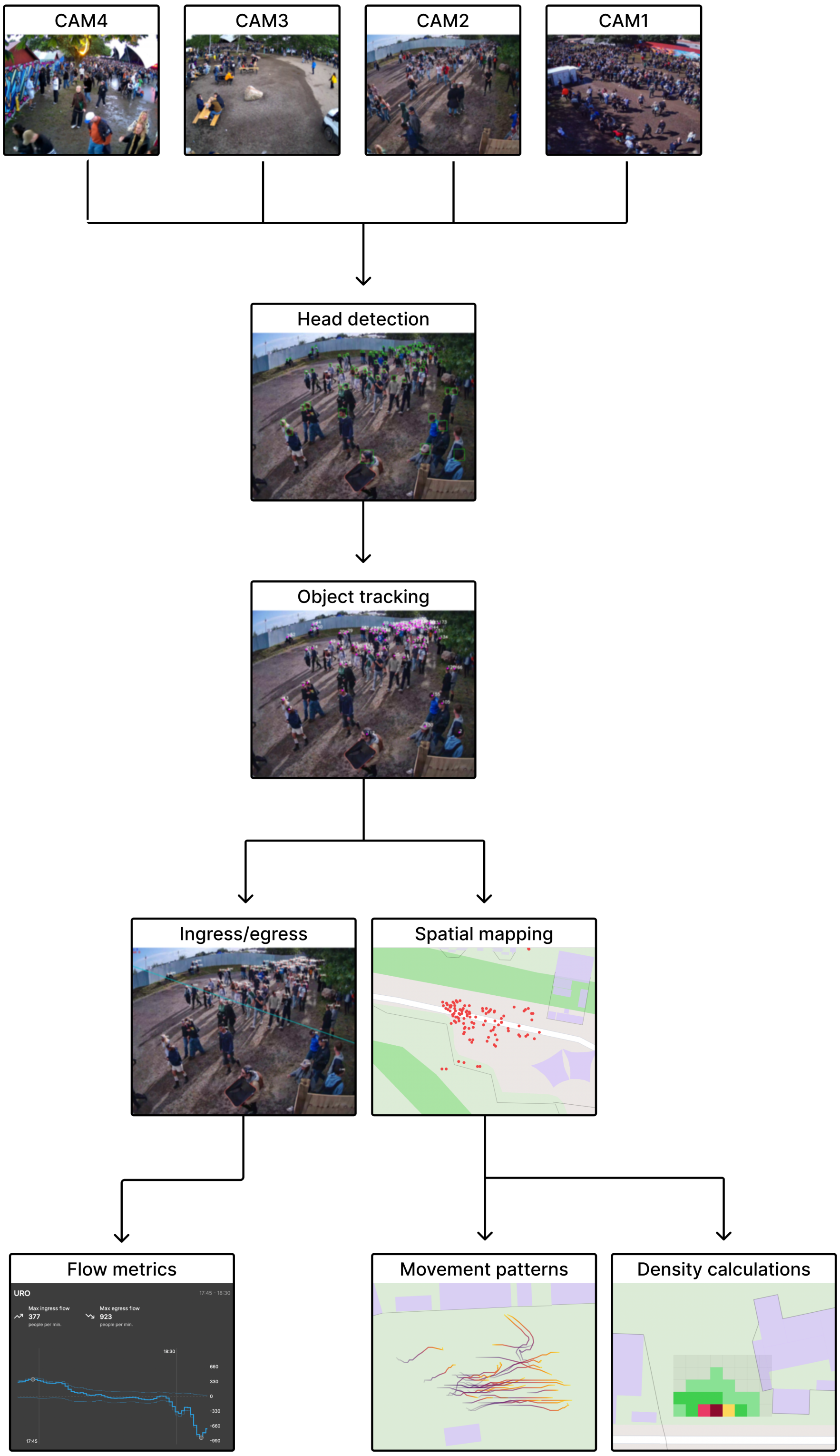


Figure 3.1: Flowchart visualizing the system architecture. *Images have been intentionally blurred to protect privacy.*

at the far eastern corner of the festival grounds, and it was anticipated that the majority of attendees would approach from the western side, where the remainder of the festival's stages are located. This involved three possible entrances/exits: *the Stables*, *the Graffiti Walk*, and *the Fast-Track*. As the Graffiti Walk is the broadest and most heavily trafficked route, *CAM1* was mounted atop a tall utility pole to ensure comprehensive coverage of this wide pathway. *CAM2* was mounted on an aluminum pole to overlook *the Fast-Track*, and *CAM4* was positioned to capture the stables entrance. Finally, *CAM3* was placed at the southeast corner of the stage, at a junction point of *the Fast-Track* and Entrances 5 & 6. This camera was positioned to capture individuals entering and exiting the latter two pathways. Altogether, these camera positions were designed to theoretically provide full coverage of the Arena stage's entrances and exits, allowing accurate metric extraction (Section 3.2.4). See Figure 3.3 for a visualization of the camera placements.

The Reolink RLC-520A cameras also included infrared (IR) night vision capabilities, theoretically allowing for monitoring in low-light conditions. However, this feature proved ineffective in practice as the integrated software lacked functionality for time-based switching between day and night modes, offering only a subjective slider for ”ambient brightness”. This created unpredictability in terms of when the cameras would switch between modes, sometimes leading to the cameras using IR imaging during daylight hours. As this resulted in reduced image quality and limitations on the video frame rate, the cameras were configured to operate without the IR functionality. Therefore, the cameras were set to record continuously for only 12 hours each day, from 12:00 to 24:00. Furthermore, due to the cameras' maximum SD card storage capacity of 256 GB, archived footage required manual download and purging when relocating the equipment between stages.

3.2.2 Computer vision model

The core of the system's ability to analyze crowd dynamics relies on a robust computer vision model capable of detecting individuals within the video footage. This section details the model selection, training process, as well as its implementation for inference.

Model selection

The primary objective of the computer vision model in this system is the detection and localization of individuals' heads within video frames; both are prerequisites for subsequent tracking and the metric extraction. Therefore, the selected model must provide image coordinates for each detected head, rather than merely an aggregate count. Methodologies in crowd analysis typically follow either density estimation approaches, which generate maps representing crowd concentration, or detection/localization-based approaches, which identify the coordinates of each individual, often via points or bounding boxes. As tracking individual trajectories is fundamental to this project's goals, localization-based methods were deemed most relevant.

The model selection process involved evaluating architectures benchmarked on an established crowd analysis dataset, NWPU-Crowd [37]. Other prominent datasets known for their complexity include ShanghaiTech and JHU-CROWD++ [38] [39]. These datasets encompass diverse scenarios and significant variations in crowd density.

One candidate architecture considered was CrowdHat, a recently published model aimed at enhancing the localization performance of standard crowd analysis models within heavily crowded environments [40]. Given its high position on the NWPU-Crowd localization leaderboard as well as its availability as open-source software, CrowdHat was selected for initial evaluation.

Another prominent architecture evaluated was You Only Look Once (YOLO), representing a widely adopted family of object detectors known for high inference speed [41]. This project employed the YOLOv8 implementation by Ultralytics [42]. Standard YOLOv8 models provided by Ultralytics are pre-trained on the large-scale COCO dataset, providing a generalized object detection capability [43]. However, achieving optimal performance for the specific domain of festival crowds necessitates fine-tuning this pre-trained model on representative video data captured during the event. YOLOv8 was chosen for comprehensive testing and ultimate deployment due to its established performance benchmarks, the ease of implementation offered by the Ultralytics library, as well as its significant advantage in processing speed.

A comparison of processing efficiency revealed that CrowdHat processes frames at an average of 261.5 ms/frame, whereas YOLOv8 achieves an average inference time of 66 ms/frame on the identical hardware configuration. Additionally, YOLOv8 demonstrated superior accuracy to CrowdHat, as seen in Table **??**. Considering the practical requirement for efficient video analysis alongside robust detection accuracy, YOLOv8 offered a superior balance. Consequently, the fine-tuned YOLOv8 model was selected as the definitive detector for integration into the system.

Model	Proc. Time (ms/frame)	Avg. Precision (%)	Avg. Recall (%)	Avg. F1-Score (%)
CrowdHat	261.5	31.70	74.26	42.32
YOLOv8	66	87.07	70.92	77.45

Table 3.1: Comparison of processing speed and model performance between CrowdHat and YOLOv8 on footage from Roskilde Festival. CrowdHat provides a slightly improved recall, but underperforms in all other metrics, as reflected in the F1-Score and processing speed.

Annotation

While pre-training on large, diverse datasets like COCO provides the YOLOv8 model with a robust general object detection capability, achieving optimal performance necessitates fine-tuning on a custom-annotated dataset. The rationale for this extends beyond simply adapting to the general festival environment; the goal is to develop highly specialized models optimized for the precise conditions and appearance characteristics encountered by each camera at each specific deployment position. It is hypothesized that such hyper-specialization yields superior detection accuracy compared to a more generalized model.

The annotation process utilized Label Studio, an open-source data labeling tool [44]. A self-hosted instance was employed to ensure data privacy and control, preventing the need to upload potentially sensitive video material captured on-site to third-party services.

Random frames were extracted from the video recordings detailed in Section 3.2.1. The annotation target was specifically the heads of individuals, rather than full bodies. This choice was predicated on the assumption that in dense crowd scenarios, heads are more consistently visible than entire bodies, providing a more reliable feature for detection and subsequent tracking. This is reflected in Roskilde Festival's Levels of Service model illustrated in Figure 2.1. Bounding boxes were drawn around each identifiable head, assigned the single class label ”person”.

Specific annotation guidelines were established to ensure consistency:

- Bounding boxes were drawn tightly around the visible extent of the head, explicitly including hair.
- In cases of occlusion, where one head partially blocks the view of another, overlapping bounding boxes were permitted.
- If individuals were distant (e.g., in the far background) or within extremely dense parts of the crowd, making distinct heads difficult to discern, bounding boxes were only placed where a head could be clearly distinguished. Ambiguous cases were omitted to maintain label quality.

To minimize time spent annotating, an iterative approach was used, as illustrated in Figure 3.4. After an initial batch of frames was annotated, a YOLOv8 model was trained on this preliminary data. This temporary model was then integrated with Label Studio to pre-annotate subsequent frames by suggesting bounding boxes based on its predictions. This workflow is commonly referred to as human-in-the-loop, and significantly expedited annotation time, as the task increasingly involved refining or validating the model's suggestions rather than manually creating annotations. Additionally, this methodology offered a visual indicator of model performance, as the diminishing need for manual adjustments indicated that the model was becoming more robust and that enough data had been annotated for final training.

The annotation results are summarized in Table 3.2.

Stage	Camera	Images	Bounding Boxes
Eos	CAM1	N/A	N/A
	CAM2	46	5297
	CAM3	N/A	N/A
	CAM4	18	1933
Arena	CAM1	22	2480
	CAM2	32	1269
	CAM3	26	2075
	CAM4	N/A	N/A
Total		**144**	**8774**

Table 3.2: Annotation statistics per camera deployment. Note that *CAM1* and *CAM3* at Eos stage were not annotated, as they were redundant to *CAM2* and *CAM4* respectively. *CAM4* at Arena stage was also not annotated due to time constraints.

Training

To develop the specialized models required for each camera deployment, the YOLOv8 model, pre-trained on the COCO dataset, was fine-tuned using the custom annotated dataset (detailed in previous section). This process utilized the Ultralytics framework [42].

Training was configured with the following hyperparameters: an input image size of 1280x1280 pixels, and training duration of 600 iterations, or epochs, with early stopping patience of 200 epochs, after which training will be terminated if no model improvement is observed. The process was optimized for the single ”person” class. To accelerate the computationally intensive process, training was performed on a computer equipped with an Nvidia RTX 4080 SUPER graphics card, leveraging GPU acceleration.

As is standard in training machine learning models, the annotated images were divided into three portions for each camera view: training, validation, and test sets. The training set comprises the majority of the data, in this case 80%, and is directly used to train the model. The validation set is used to evaluate the model's performance on unseen data during each iteration, testing the model's ability to generalize its learning. The final portion of the dataset, the test set, is reserved for final evaluation of the model's performance after training is complete. The test set is not used during training or validation, ensuring that the model's performance is assessed on completely unseen data.

This fine-tuning stage produced specialized model weights adapted to the unique visual characteristics of each camera view at the festival, which were then used for the subsequent inference stage.

Inference and Tracking

Following the training phase, the inference stage employs the specialized YOLOv8 models to detect heads within the recorded video footage, which are then tracked across frames using an object tracking algorithm. This combined inference and tracking pipeline generates the foundational data required for subsequent spatial mapping and metric extraction. The process begins by loading the fine-tuned model specific to the camera view being analyzed. Input videos are processed in parallel, analyzing each frame individually to identify and track individuals present. Videos are processed at a rate of 15 frames per second (FPS), which is half the recorded frame rate of 30 FPS. This frame rate was selected to balance processing speed and accuracy, as experimentation showed that tracking performance was not significantly affected by halving the frame rate. In theory, this provides a 2x speedup in processing time.

Each frame is resized to a 1280x1280 pixel resolution, as defined during training, before running inference. This prepared frame is then passed to the fitted YOLOv8 model, which outputs bounding boxes around predicted heads and assigns confidence scores to these detections. These predictions are filtered; detections falling below a predefined confidence threshold are discarded (40%), and Non-Maximum Suppression (NMS) is applied to resolve significant overlaps between bounding boxes, preserving only the most confident prediction.

The resulting bounding boxes and confidence scores are subsequently passed as input to the object tracking module. This project utilizes ByteTrack, a high-performance algorithm chosen for its accuracy, particularly within crowded scenes [45]. ByteTrack associates the current frame's detections with previous frames, assigning a unique ID to each tracked individual.

The final output for each frame is a list containing the bounding box coordinates, the assigned tracking ID, and the detection confidence score for every tracked individual. This data serves as input for the spatial mapping system described in the following section.

3.2.3 Spatial mapping and GIS

While the computer vision model (Section 3.2.2) outputs the locations of individuals in terms of pixel coordinates within the video frame, these coordinates alone are insufficient for thorough analysis of crowd dynamics. In order to derive area-based metrics such as crowd density (people per square meter), movement speeds and distances, it is necessary to translate these pixel positions into real-world geographic coordinates. Spatial mapping has this purpose, providing the planar transformation between the camera's perspective and the geographic context of the festival. This mapping also enables visualization of individual positions, from multiple cameras, onto a single overhead map, significantly enhancing contextual understanding, compared to that which is achievable through video footage alone.

The technique employed for this spatial mapping is *homography*: a transformation that maps points from one plane to another. In this context, it establishes a mathematical translation between the pixel coordinates in the 2D camera image plane and the corresponding real-world coordinates on the ground plane [46]. This allows any detected pixel coordinate within a defined area to be projected onto its actual geographic location.

Performing the homography calculation requires two sets of corresponding points: one set in the camera's pixel coordinates and another in real-world coordinates. The latter is obtained from Roskilde Festival's internal Geographic Information System (GIS) tooling, which provides precise GPS coordinates of all infrastructure on the festival grounds. Selecting corresponding points was achieved by identifying distinct, stationary landmarks visible in the video frame, such as corners of structures or fences, and marking their position in pixel coordinates. These landmarks were then located in the GIS utility, where their GPS coordinates were recorded. Given a minimum of four distinct pairs of corresponding points, the homography transformation can be computed – in this case, utilizing the OpenCV Python library [47]. This resulting transformation is stored in a configuration file associated with that camera deployment. This manual process was performed once for each camera deployment, and the resulting homography matrices were used for all subsequent video footage captured by that camera. See Figure 3.5 for an illustration of the results of this mapping process.

Note a slight limitation of homography in this context; homography assumes that the mapping occurs between two planar surfaces. While the surfaces in the camera and map are treated as planar, the Earth's surface is curved. For the relatively small geographic areas covered by individual camera views, the ground surface can be reasonably approximated as flat. The error introduced by this assumption is considered negligible for the purposes of crowd analysis at this scale.

3.2.4 Metric extraction

Following the spatial mapping process (Section 3.2.3), which translates tracked individuals' pixel coordinates into real-world geographic coordinates, the metric extraction system processes this positional data to derive quantitative insights into crowd dynamics. This stage is the crucial step transforming raw data into the key metrics utilized by crowd safety professionals. The primary metrics extracted, as defined in Section 2.1.2, include ingress/egress counts, flow rates, cumulative counts, crowd density, and movement patterns.

Ingress/Egress and Flow Rate Calculation

To measure the flow of people into and out of specific areas, lines are drawn across each camera view. These lines correspond to the entrances/exits covered by the cameras (as detailed in Section 3.2.1, Figures 3.2 and 3.3). The system follows the trajectory of each tracked individual (identified by a unique ID from the tracking algorithm, Section 3.2.2). When a trajectory crosses the virtual line, the system registers it as either an ingress or egress event based on the direction of crossing.

These individual crossing events are then aggregated over specific time intervals. Intervals of 1 minute, 15 minutes, and 1 hour were identified as most useful by Roskilde Festival (Appendix A.2.6). This aggregation allows the calculation of the ingress/egress flow rate, representing the number of people entering/exiting the area per time unit. The net flow rate is calculated as the difference between ingress and egress flow rates, indicating the rate of change in the number of people within the area.

Furthermore, a cumulative count provides a running total of the net number of people within the monitored area over time, calculated by aggregating ingress/egress over time. This cumulative count is particularly useful for understanding the overall occupancy of the area. These flow metrics can be calculated for individual entrances or cameras, or they can be aggregated to provide a total flow for a larger area such as an entire stage. The system also identifies and records maximum ingress and egress flow rates observed during specific periods, like during a concert. This data is essential for understanding peak loads, capacity planning, and validating entrance width calculations.

Crowd Density Calculation
Crowd density, measured in people per square meter (people/m²), is another critical metric for assessing safety and comfort. To calculate density, the monitored area is divided into a grid of 3x3 meter cells. At discrete time intervals, the system counts the number of tracked individuals whose mapped geographic coordinates (Section 3.2.3) fall within each grid cell. The density for that cell or zone is then calculated by dividing this count by the known area of the cell or zone.

Movement Patterns
Beyond counts and density calculations, understanding how crowds move is crucial. By analyzing the sequence of time-stamped geographic coordinates, or trajectories, associated with each unique tracking ID, the system can visualize movement patterns. This involves plotting the paths taken by individuals over time, represented as gradient lines on the map.

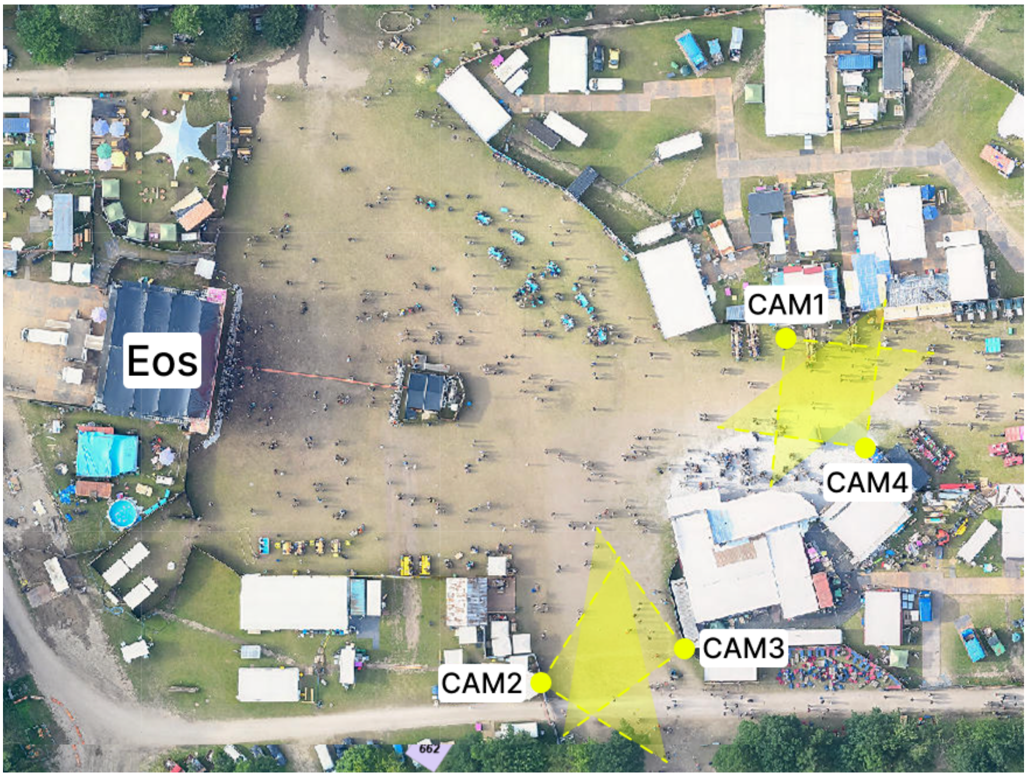
Analyzing these patterns can reveal dominant flow directions, identifying the primary paths people take when moving between locations. It can also aid in detecting cross-flow, areas where different streams of people intersect. Additionally, this analysis helps understand origin-destination patterns, showing where people come from when approaching an area, such as which pathway they used, and where they head afterwards.

3.2.5 Interface/frontend

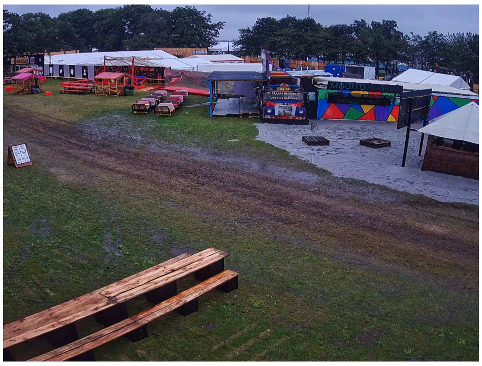
The interface/frontend is the final component of the system, and is the only component visible to the end-users. It is therefore designed to condense the complexity of the subsystems described in the previous sections into an accessible and user-friendly interface. The application employs a modern web development stack: namely, React, Next.js, TypeScript, and Tailwind CSS.

Data storage and retrieval is managed through a PostgreSQL database, accessed via the Prisma Object-Relational Mapper (ORM). The database contains predefined information, including camera configurations, processed count data, and timestamped geographic point data. Additionally, the database stores user login information, as well as user-generated labels/notes for specific time intervals and cameras.

The entire frontend application is deployed and hosted using Vercel. See full overview of the frontend application in Section 4.1.



(a) Eos stage area – visualized with camera placements



(b) CAM1 Preview

(c) CAM2 Preview

(d) CAM3 Preview

(e) CAM4 Preview

Figure 3.2: Camera placement at Eos stage, with approximate field of view (FOV) indicated (a). The bottom images show sample frames from the four cameras deployed at the Eos stage, showing the field of view for each camera position (b-e).

(a) Arena stage area – visualized with camera placements

(b) CAM1 Preview

(c) CAM2 Preview

(d) CAM3 Preview

(e) CAM4 Preview

Figure 3.3: Camera placement at Arena stage, with approximate field of view (FOV) indicated (a). The bottom images show sample frames from the four cameras deployed at the Arena stage, showing the field of view for each camera position (b-e).

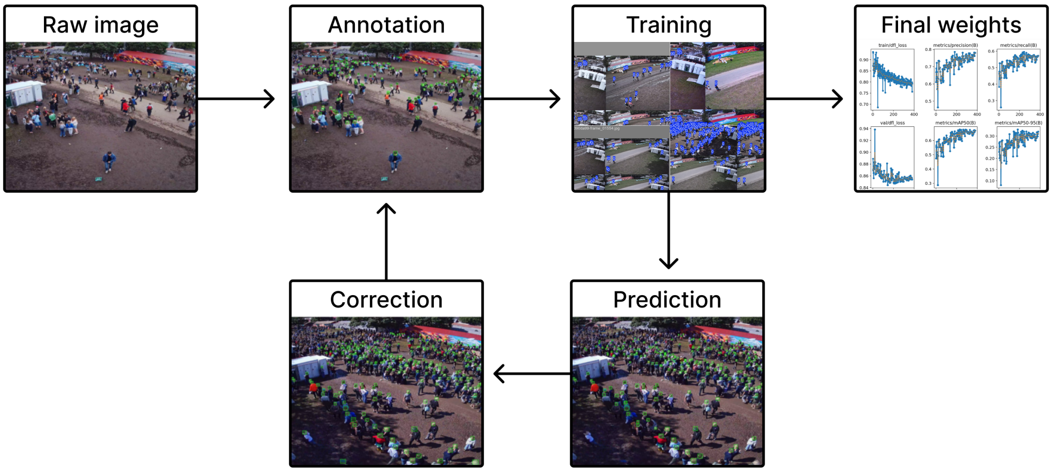


Figure 3.4: Flowchart visualizing the annotation process. *Images have been intentionally blurred to protect privacy*.

(a) Camera view

(b) Map view

Figure 3.5: Illustration of homography mapping between camera view and map view.

4 Build-Measure-Learn

The following chapter presents the results of the project, as well as serving as a summary of the process dictated by Eric Ries' Build-Measure-Learn feedback loop [11]. The chapter is divided into three main sections: one for each stage of the iteration process. The Build section showcases the resulting frontend application, or essentially the product of the project's final *build* iteration. The Measure section presents a technical performance evaluation, as well as the results of the workshop held with the Roskilde Festival safety team. The latter also serves as a representation of the preceding *measure* phases, where the product was presented to the safety team for feedback. Finally, the Learn section reflects on the value of the product, as well as assessing its fulfillment of the requirements outlined in previous chapters, representative of the evaluation conducted in each *learn* phase.

4.1 Build

Following the Build-Measure-Learn framework, the application was developed iteratively, with each version being refined based on user feedback. The process started with an extended Learn phase, where many conversations were held with the safety team in order to thoroughly grasp their needs and existing workflows. The product's first version was then built (Build), and Roskilde Festival's Mads and Morten Therkildsen were subsequently invited to provide feedback. Feedback in this case is the qualitative data netted from the Measure phase. This feedback was then analyzed and used to define new features and refine the product, which was then again presented and evaluated. The process was repeated until the final version was reached, which was presented in the workshop outlined in Section 4.2.2.

This section showcases the resulting frontend application, which serves as the primary interface for the Roskilde Festival safety team to access and analyze crowd dynamics data. The interface is designed to be intuitive, presenting complex data through interactive charts and maps. The following figures illustrate the key features and functionalities of the application. Note that data shown from June 30th to July 2nd corresponds to the Eos stage, while data from July 3rd onwards corresponds to the Arena stage.

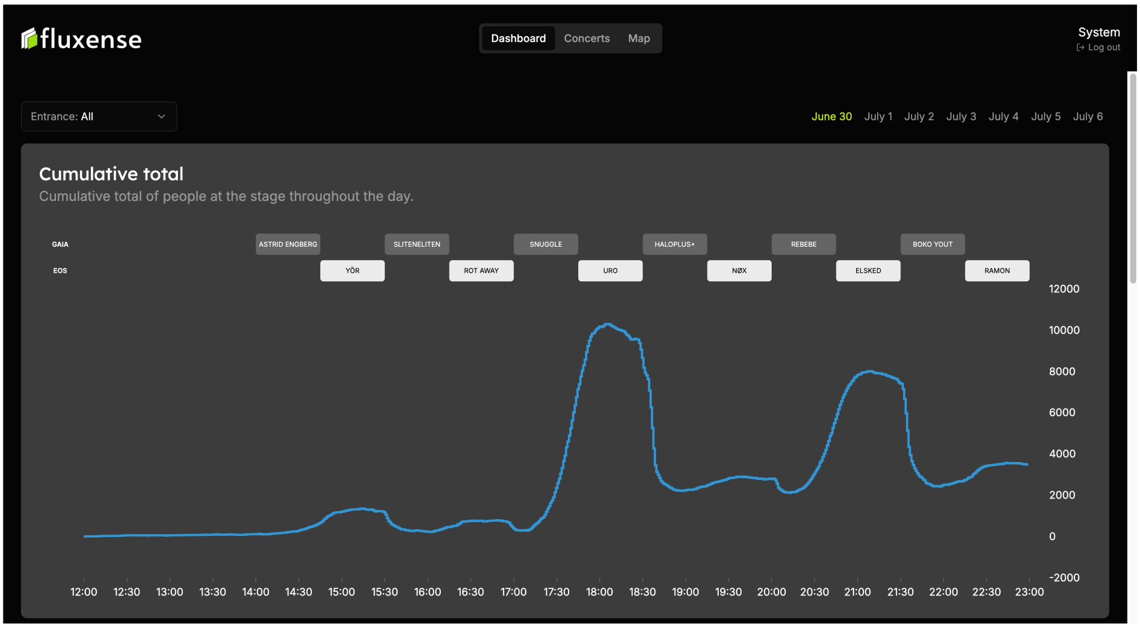


Figure 4.1: The 'Dashboard' view displaying the cumulative total number of people estimated to be at the selected stage (Eos or Arena) throughout the day. An overlay along the top indicates the scheduled concert timings for the observed stage. Based on feedback requesting analysis of inter-stage flow (Appendix A.2.5), schedules for adjacent stages (i.e., Gaia when observing Eos, or Orange stage when observing Arena) were also included.

Figure 4.2: The 'Dashboard' view showing the ingress (positive values) and egress (negative values) rates in people per minute throughout the day. Displaying data split by individual entrances (e.g., "Entr. 10", "Tow. Gaia") was implemented based on feedback requesting insight into the load of each entrance. The concert schedule overlay, including schedules from relevant adjacent stages added per user feedback, is also present for context.

Figure 4.3: The 'Concerts' view presenting the overall flow rates. It displays ingress-flow (top line), egress-flow (bottom line), and the total flow (bold middle line). Based on feedback, flow rates are presented in people per minute for easier interpretation compared to the initial people/second unit (Appendix A.2.6). Hovering over the chart reveals a tooltip displaying instantaneous ingress, egress, and total flow rates at a given minute, along with aggregated ingress/egress totals over the last 15 and 60 minutes, aligning with RF's requested analysis intervals. The concert schedule overlay is also present for context.

Figure 4.4: The 'Concerts' view, with a breakdown of the flow rate data for individual concerts scheduled on the selected day. Each concert card shows the flow during the concert period and highlights the maximum observed ingress and egress flow rates in people per minute.

Figure 4.5: The filtering controls available in the application, refined based on user requirements identified during feedback sessions. Users can select the date and filter the displayed data by specific entrances or view aggregated data for all entrances at a stage.

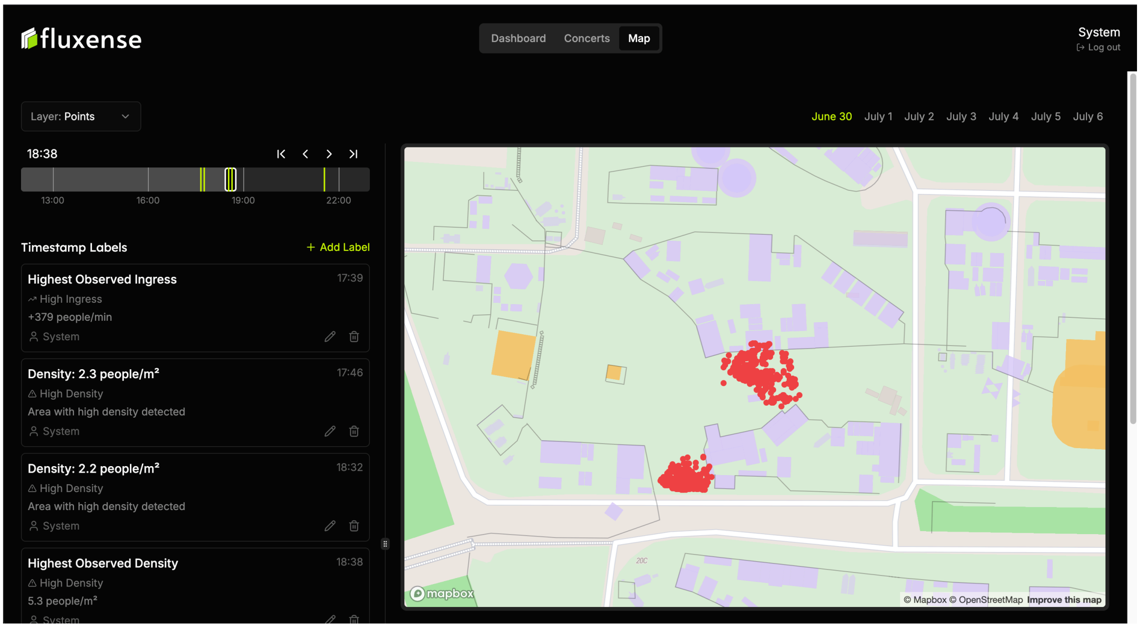


Figure 4.6: The 'Map' view interface elements. This includes an interactive timeline slider for navigating through the day, the 'Timestamp Labels' feature for adding notes, and the map displaying positional data for the selected time. The map layer itself can be configured to show individual detections as scatter points, as seen here.

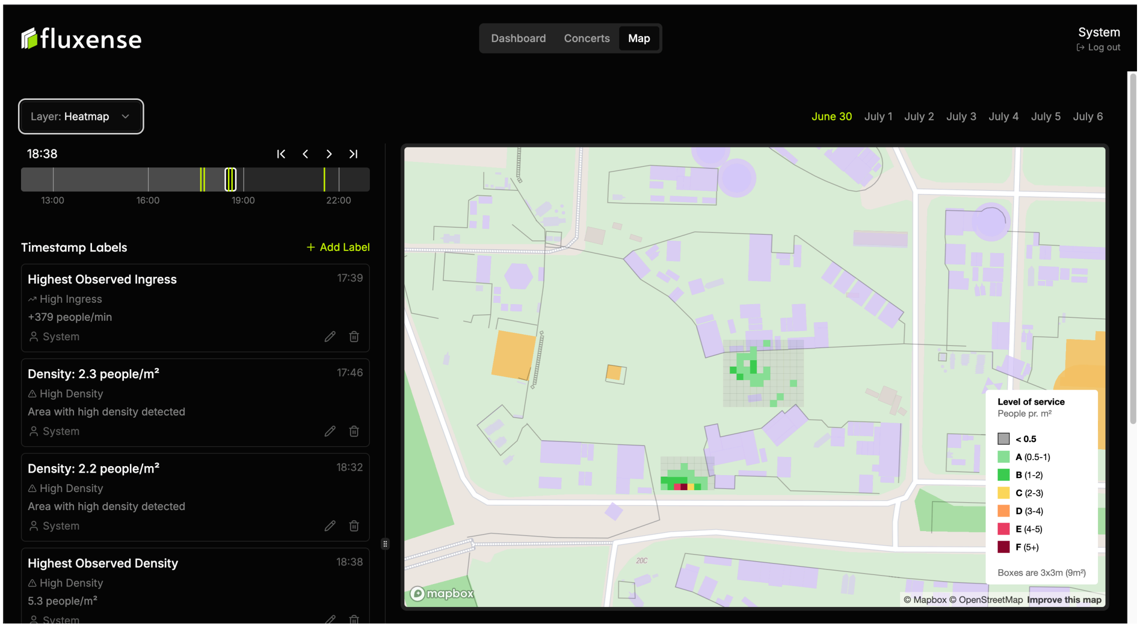


Figure 4.7: The 'Map' view displaying the 'Heatmap' layer. Crowd density is visualized using colored 3x3 meter grid cells. Following feedback, the color scale corresponds to Roskilde Festival's internal Levels of Service (LoS) scale (A-F), indicating people per square meter (people/m^2) to align with their existing practices.

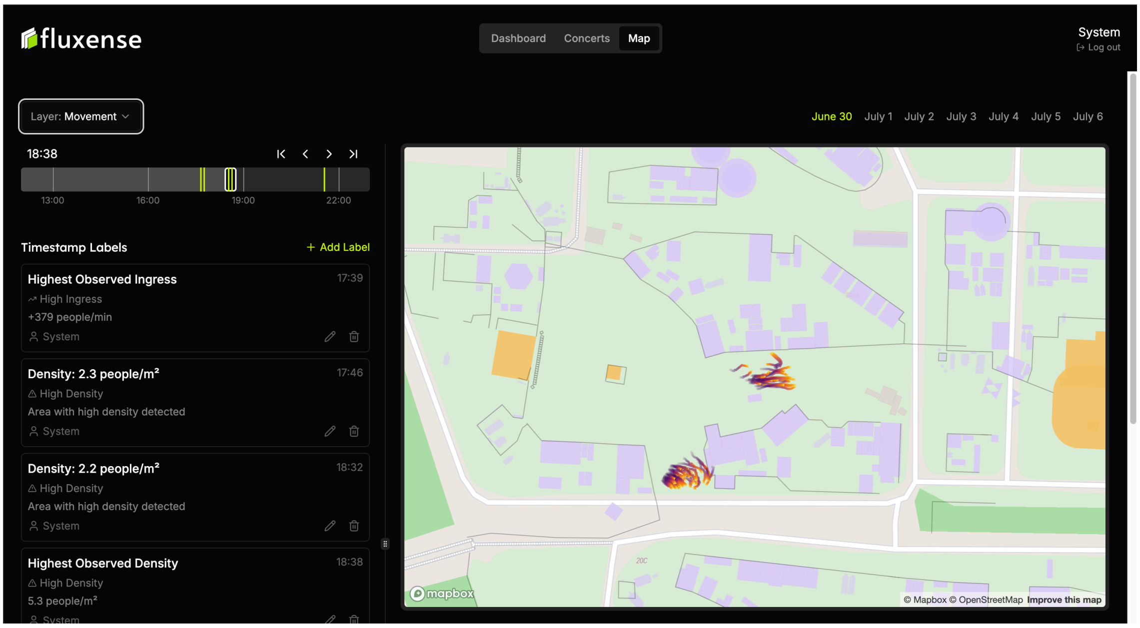


Figure 4.8: The 'Map' view displaying the 'Movement' layer. Developed in response to feedback requesting visualization of movement patterns, it shows the trajectories of tracked individuals as lines on the map, with color gradients indicating the direction of movement, helping to visualize flow and origin-destination patterns.

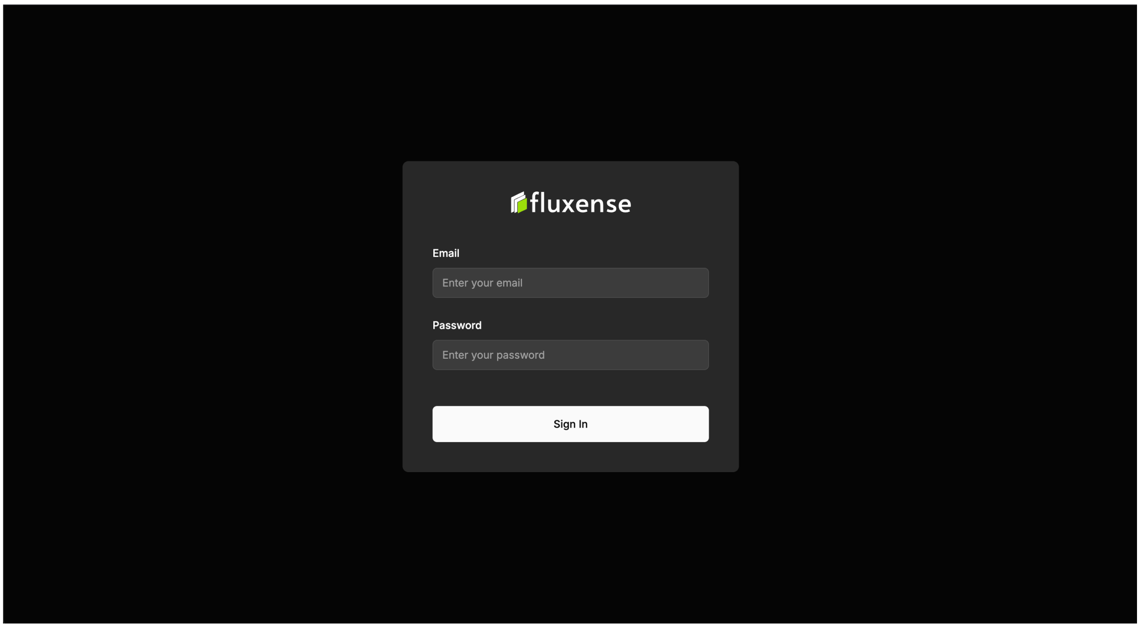


Figure 4.9: The application login screen, ensuring secure access to the crowd analysis data.

4.2 Measure

This section presents an evaluation of the technical performance of the developed product, as well as the value of the product as perceived by the Roskilde Festival safety team. The evaluation was conducted through a workshop with key members of the safety team, where they were presented with the product and asked to provide feedback on its performance and value as compared to their existing workflows.

4.2.1 Technical performance

The computer vision model selected for detecting individuals in video footage was YOLOv8, which was subsequently fine-tuned for the specific task of head detection. This specialized model is further referred to as *yolo_heads*. The rationale for selecting YOLOv8 included its balance of high inference speed and robust detection accuracy, making it suitable for processing extensive video data efficiently (as noted in Section 3.2.2). To optimize performance for the unique conditions of each camera's viewpoint and the general festival environment, separate *yolo_heads* models were trained using custom annotated frames extracted from video recordings from each specific camera deployment at the Eos and Arena stages.

The performance of the fine-tuned *yolo_heads* models for the key camera deployments used in this study was evaluated using standard object detection metrics: Precision, Recall, and F1-score. Precision measures the accuracy of the positive predictions (i.e., what proportion of detected heads were actual heads). Recall measures the model's ability to find all actual heads (i.e., what proportion of actual heads in the frames were detected). The F1-score provides a harmonic mean of Precision and Recall, offering a single measure of the model's accuracy. It thus symmetrically represents both precision and recall in one metric. As is standard in evaluation of machine learning models, a separate test set of annotations was used to evaluate the model's performance. This ensures that the performance assessment is not biased by the data used for training the model.

The results for the *yolo_heads* models on footage from Arena's *CAM1*,*CAM2*, and *CAM2*, as well as Eos' *CAM2* and *CAM4* are presented in Table 4.1. Note that *CAM1* and *CAM2* at Eos stage were not utilized, as they were redundant to *CAM2* and *CAM4* respectively. *CAM4* at Arena stage was also not utilized due to time constraints.

Stage	Camera	Precision (%)	Recall (%)	F1-Score (%)
Eos	CAM2	91.23	76.52	83.23
	CAM4	86.94	81.84	84.32
Arena	CAM1	79.92	41.25	54.42
	CAM2	92.62	84.01	88.11
	CAM3	84.62	70.97	77.19
Average		**87.07**	**70.92**	**77.45**

Table 4.1: Performance metrics of the fine-tuned YOLOv8 head detection model across different camera deployments at Roskilde Festival.

4.2.2 Roskilde Festival workshop

In order to evaluate the business value of the developed product, a workshop was held with two key members of the Roskilde Festival safety team, Mads Therkilden (Safety Lead) and Niels Laustsen (Program Safety Team Lead). The workshop began with a demonstration of the results of the project, where participants were shown the final product and its capabilities, subsequently providing login credentials to access the application themselves. The remainder was structured around a series of scenarios that the safety team might encounter during the planning of the festival. The scenarios were designed in order to test whether the developed product could provide a concrete value in certain situations. The scenarios were as follows:

1. **Scenario 1:** Entrance 10 is a primary access point for the Eos stage, especially crucial during the first days. Let's assume festival-goer feedback from Roskilde Festival (RF) 2024 suggested this entrance felt congested at times. The safety team needs to assess if the current width is adequate or requires modification for RF 2025 to handle peak demand safely.
2. **Scenario 2:** Your colleague from Food & Beverage (F&B) proposes moving the popular Cava Bar to the main corridor leading to/from Entrance 10 to capture more foot traffic. The safety team is concerned because they believe this corridor already experiences some congestion during peak times. Adding a bar with likely queues could create unsafe scenarios. The task is to explain to the F&B colleague these concerns and justify recommending against the relocation.
3. **Scenario 3:** A colleague questions the back-to-back scheduling of the Eos and Gaia stages during the first days, concerned that it could lead to congestion between the stages. How can you determine and communicate whether or not the scheduling strategy was operationally sound?

After presenting each scenario, the participants were requested to discuss how they would approach the scenario given their current workflows. Once they reached an agreement, they were asked to use the application to aid in solving the same scenario. After each the scenario, the participants were presented with a series of questions, aimed at quantitatively evaluating the business value of the product. The results of this evaluation can be found in (Appendix A.1). The questions were as follows:

1. Overall, compared to your current process, how valuable do you perceive the approach using this tool to be for this specific task? (*Scale 1-5: Not at all Valuable to Extremely Valuable*)
2. How do you think using this tool might impact the time or effort required for this task compared to your current methods? (*Scale 1-5: Much More Effort to Much Less Effort*)
3. Compared to your current process, how might using the tool affect your confidence in the outcome related to this task? (*Scale 1-5: Much Less Confident to Much More Confident*)

Assessing an issue such the first scenario, assessing Entrance 10, is challenging in their current workflows, as it heavily depends on the availability of relevant CCTV footage, which is not guaranteed. Assuming footage exists, the team would need to manually search through it, and make subjective density estimates based on the Levels of Service model presented in Figure 2.1. Given access to the application, the team could immediately find the historical density data for the specific entrance, identify the peak moments (e.g., a brief 5.3 people/m^2 spike), and observe a dissipation after only a few minutes. This aided the team in concluding that the short-lived peak was not concerning enough to warrant a significant change in the existing infrastructure.

The second scenario, concerning the Cava Bar relocation, clearly contrasted current communication efforts. Assuming the team has access to the information gathered in the first scenario, the current workflow relies heavily on verbal arguments to justify safety concerns to other departments, such as Food & Beverage. If further justification is needed, Mads Therkildsen explained he would manually create visuals, for instance using PowerPoint, to illustrate theoretical concerns like cross-flow or queue formation – a time-consuming process requiring skills outside their core competencies. Given the application, the team could instantly collect objective visual evidence by taking screenshots of density heatmaps or movement flow patterns.

Finally, the third scenario, involved back-to-back concert scheduling between Eos and Gaia. The current workflow bases planning on estimations: calculating required pathway widths using the theoretical maximum capacity of a stage (e.g., 8000 for Eos) and applying standard formulas based on pathway width. This method relies on assumptions about crowd size, which is currently impossible to accurately calculate, as well as expected behavior based on concert assessments, as discussed in Section **??**. The safety team subsequently used the application to identify actual peak attendance (closer to 10,000) and, more importantly, the actual peak flow rate through the specific connecting pathway (around 350 people/minute). This information allowed the team to rapidly assess that the existing infrastructure was sufficient to support maximum capacity.

Following the three scenarios, the participants were asked to reflect on the overall value of the product, as well as brainstorm use-cases that were not initially anticipated. The team found the greatest potential emerging given more footage and data across the festival site. Mads Therkildsen expressed that the product could play an integral role in planning entrances around the Arena stage, which is a chronic concern for the safety team. Arena is located in the eastern corner of the festival site, contributing to a significant imbalance in the distribution of attendees. Several attempts have been made to address this issue, but Mads sees potential for the product to provide objective data to evaluate their effectiveness in distrusting festival-goers. Niels Laustsen provided a similar example, where the product could be used to determine fencing layout in front of the Orange Stage. He directly refers to the Astroworld incident mentioned in the introduction (Section 1.1), and ideated that the product could be used to determine ingress imbalance around the front pit area. Additionally, Niels expressed a frustration with the current process of communicating safety requirements, and stated that the product could likely save him around two days of work in the planning phase, and that is just from this application alone.

4.3 Learn

The following section reflects on the value of the product, as well as assessing its fulfillment of the requirements outlined in previous chapters.

4.3.1 Requirement fulfillment

The resulting frontend application directly addresses the requirement specifications outlined in Section 2.3.

Core metric extraction (Requirement 1) is achieved through the backend processing pipeline (detailed in Chapter 3), with the frontend serving as the interface for these metrics. Figures 4.1 through 4.4 demonstrate the visualization of key metrics including cumulative totals, ingress/egress, flow rates, and concert-specific flow analyses, with movement patterns and crowd density heatmaps visualized in Figures 4.8 and 4.7, respectively.

Additionally, the platform provides an intuitive user interface (Requirement 2), by including interactivity through filtering (Figure 4.5), and visualization tooltips (Figure 4.3), ensuring that users can easily filter and understand the presented data. This was demonstrated during the workshop held with Roskilde Festival, where users were able to intuitively navigate the platform following a brief demonstration. Given the results of the workshop (Section 4.2.2), the users were evidently able to quickly access and interpret the data, as well as utilize the platform to solve the scenarios presented.

Regarding compliant data handling (Requirement 3), the system is designed for compliance with data protection regulations (Section 2.4). Most importantly, no biometric or personally identifiable data is stored by the platform. Immediately following its input into the object detection model (Section 3.2.2), video footage is discarded. The images are converted into quantitative and geographic positional data, ensuring that all historical data retained is fully anonymized, thereby protecting privacy while still providing valuable insights. In order to further secure the data provided by the platform, the application is protected by a login screen (Figure 4.9), ensuring that only authorized persons from Roskilde Festival can access the site.

4.3.2 Evaluating value

The workshop with key members of the Roskilde Festival safety team (Section 4.2.2) served as a crucial step in demonstrating the product's value by applying it to realistic situations faced by the safety team. Analysis of the discussions and feedback for each scenario reveals how the tool addresses the specific objectives defined in Section 1.5, often by providing objective data lacking in current workflows.

Objective 1: Improve internal communication

The tool's ability to improve internal communication was evident in scenarios like the Cava Bar relocation (Scenario 2) and Entrance 10 assessment (Scenario 1). In both cases, the application provided immediate access to objective visual evidence, replacing the need for verbal explanations or manually creating visuals (e.g., in PowerPoint), which is time-consuming and outside the safety team's core competencies. The significant effort reduction, rated 5/5 by both participants for these scenarios, and Niels Laustsen's comment on potentially saving "around two days of work" just for communicating safety requirements, underscore this benefit.

Objective 2: Enhance safety planning

The product demonstrated a strong potential to enhance safety planning. For instance, in the Eos and Gaia stage scheduling (Scenario 3), the tool provided actual peak attendance and flow rate data, allowing for a validated assessment of infrastructure adequacy rather than relying on estimations based on theoretical maximums. This shift towards data-driven validation increased confidence in planning decisions (rated 4/5 by both participants). Similarly, in the Cava Bar scenario (Scenario 2), data on existing congestion patterns allowed for making a proactive risk assessment, preventing a potentially unsafe bar location.

Objective 3: Create reliable documentation

The tool's capability to create reliable documentation was also clearly shown. When assessing Entrance 10 congestion (Scenario 1), the application offered immediate access to historical density data. This objective record replaces time-consuming manual searches through potentially incomplete CCTV footage, and subsequent subjective density estimations. The high ratings for value (4/5) and effort reduction (5/5) in this scenario highlight this. The data captured for stage scheduling (Scenario 3), such as actual flow rates and attendance, also contributes to a valuable historical dataset for future planning and continuous improvement.

To summarize, the workshop confirmed that the developed platform provides substantial business value by directly addressing the objectives above and improving upon current workflows. Generally speaking, the feedback was overwhelmingly positive, especially regarding improvements to internal and external communication, which was regarded as the most valuable aspect of the product. The product's value was demonstrated in enhancing safety planning with data replacing estimations, improving internal communication with accessible, objective visual evidence, and documenting for knowledge retention within Roskilde Festival's safety team.

5 Conclusion

5.1 Summary of results

This project set out to improve crowd safety measurability at Roskilde Festival by developing an AI-enabled video surveillance analysis platform. The research explored existing crowd safety management practices, identified key limitations in objective data acquisition, and proposed a technical solution to address these gaps.

The core of the project involved designing and implementing a system capable of processing video footage to extract actionable crowd dynamics metrics. This was achieved through a multi-stage process encompassing strategic data collection using deployed cameras, development and fine-tuning of a YOLO-based computer vision model for head detection and tracking, spatial mapping via homography to translate pixel coordinates to real-world GIS data, and subsequent extraction of key metrics such as ingress/egress counts, flow rates, cumulative population, crowd density, and movement patterns.

The developed system successfully met its defined objectives, proved through a workshop conducted with key members of the Roskilde Festival safety team. Participants rated the tool highly for perceived value, effort reduction, and confidence enhancement in addressing typical planning and communication scenarios. The feedback highlighted the tool's potential to streamline workflows, particularly in justifying safety requirements and in planning based on objective data rather than solely estimations or past anecdotal knowledge. The safety team also identified further potential use-cases, such as optimizing fencing layouts and evaluating the effectiveness of crowd distribution strategies.

The technical performance evaluation of the fine-tuned head detection models showed an average F1-score of 77.45% across various camera deployments, with individual models achieving F1-scores as high as 88.11% (Arena CAM2). This indicates an acceptable level of accuracy in the core detection task.

The system was designed with data compliance in mind, ensuring that video footage is processed to yield anonymized, aggregated data for storage and presentation, thereby adhering to GDPR principles.

In essence, the project successfully developed and validated a functional prototype of an AI-driven platform that provides objective, measurable insights into crowd dynamics, demonstrating clear benefits for Roskilde Festival's safety management practices.

5.2 Market expansion opportunities

While this thesis project was specifically tailored to the needs of Roskilde Festival and not intended for commercial use, the underlying technology and the insights gained have significant potential for market expansion beyond this initial application. This project has essentially created a template for other safety-focused event managers. With minor adjustments, the developed system can be adapted to other music festivals, sporting events, or even public gatherings. Where there are cameras, there is a way.

The technology can also be applied to other sectors, such as transportation hubs (airports, train stations), urban planning (monitoring pedestrian traffic in city centers), or business intelligence (analyzing customer flow in retail environments).

5.3 Technical improvements

5.4 Closing remarks

References

[1] David Fricke. *Nine Dead At Pearl Jam Concert*. Aug. 2000. URL: https://web.archive.org/web/20180830050458/https://www.rollingstone.com/music/music-news/nine-dead-at-pearl-jam-concert-235167/.

[2] NBC News. *Officials release cause of death for Astroworld victims*. Dec. 2021. URL: https://web.archive.org/web/20240927060550/https://www.nbcnews.com/news/us-news/officials-release-cause-death-astroworld-victims-rcna9097.

[3] YOUROPE - The European Festival Association. *Roskilde Festival*. URL: https://web.archive.org/web/20241212233325/https://yourope.org/member/roskilde-festival/.

[4] Vladimir Hubka and Wolfgang Ernst Eder. *Design Science: Introduction to the needs, scope and Organization of Engineering Design Knowledge*. Springer London, 1996.

[5] Wolfgang Eder. "Why Systematic Design Engineering?" In: vol. 8. Jan. 2009. DOI: 10.1115/DETC2009-86067.

[6] Nigel Cross. *Engineering Design Methods: Strategies for Product Design*. Third Edition. John Wiley and Sons, Ltd, 2000.

[7] Michael Joseph French. *Conceptual Design for Engineers*. Second Edition. Springer-Verlag Berlin Heidelberg GmbH, 1985.

[8] Gerhard Pahl et al. *Engineering Design: A Systematic Approach*. Third Edition. Springer, 2007.

[9] Karl T. Ulrich, Steven D. Eppinger, and Maria C. Yang. *Product Design and Development*. Seventh Edition. McGraw-Hill Education, 2020.

[10] Ken Schwaber and Jeff Sutherland. *Scrum Guide*. Scrumguides.org. Nov. 2020. URL: https://scrumguides.org/scrum-guide.html.

[11] Eric Ries. *The Lean Startup: How Toda's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown Business, 2011.

[12] Kulturministeriet. *Vejledning om sikkerhed ved udendørs musikarrangementer o.lign.* URL: https://web.archive.org/web/20211229000406/https://kum.dk/fileadmin/_kum/5_Publikationer/2014/Vejledning_om_sikkerhed_ved_udendoers_musikarrangementer_2014.pdf.

[13] POLITI. *Vejledning om udarbejdelse af sikkerhedsplan for større udendørs musikarrangementer o.lign.* URL: https://web.archive.org/web/20240927194102/https://politi.dk/-/media/mediefiler/landsdaekkende-dokumenter/tilladelser/begivenheder/vejledning-om-udarbejdelse-af-sikkerhedsplan-for-strre-udendrs-musikarrangementer-ol.pdf.

[14] *The Purple Guide to Health, Safety and Welfare at Outdoor Events*. Food, Events and Things (FEAT) Ltd, 2021.

[15] G. Keith Still. *Applied Crowd Science*. CRC Press, 2022.

[16] *The Event Safety Guide*. Event Safety Alliance of USA, Inc., 2016.

[17] G. Keith Still. "Ramp analysis". In: *Applied Crowd Science* (Dec. 2021), pp. 103–117. DOI: 10.1201/9781351053068-5.

[18] John J. Fruin. PhD thesis. 1970.

[19] Chris Zimmerman et al. "Space vs. Place: Comparing Space-based Movements and Place-based Experiences at the Roskilde Festival 2015". In: vol. 11. 2016, pp. 1–10. DOI: 10.1145/2930971.2930983.

[20] Apple App Store. *Smukfest 2025*. URL: https://web.archive.org/web/20250408172430/https://apps.apple.com/dk/app/smukfest-2025/id6447997155.
[21] Reddit. *Tomorrowland Map*. URL: https://web.archive.org/web/20220709110843/https://preview.redd.it/m5jhl2mvyia91.jpg?width=1170&format=pjpg&auto=webp&s=2c44480b31448b153fc81a2dc11269662ebdd985.
[22] 3 SIDED CUBE. *An App To Truly Elevate The Festival Experience For Every Attendee*. URL: https://web.archive.org/web/20250207080506/https://3sidedcube.com/projects/an-app-to-truly-elevate-the-festival-experience-for-every-attendee.
[23] GPS.gov. *GPS Accuracy*. URL: https://web.archive.org/web/20250421225547/https://www.gps.gov/systems/gps/performance/accuracy/.
[24] Wiliot. *Bluetooth Beacon*. URL: https://web.archive.org/web/20240529070543/https://www.wiliot.com/bluetooth-beacon#applications-of-bluetooth-beacons.
[25] Pietro Centorrino et al. "Managing Crowded Museums: Visitors Flow Measurement, Analysis, Modeling, and Optimization". In: *Journal of Computational Science* 53 (July 2021). URL: https://arxiv.org/abs/2006.16830.
[26] Sakurako Tanida et al. "Quantifying crowding perception at large events using beacons and surveys". In: *Scientific Reports* 15.1 (Jan. 2025). DOI: 10.1038/s41598-024-83065-4.
[27] Remark AI. *Video Intelligence for the Smart Stadium*. URL: https://web.archive.org/web/20240822104225/https://remarkvision.com/smart-stadiums-and-events/.
[28] Lumeo. *Solutions*. URL: https://web.archive.org/web/20230605023051/https://lumeo.com/solutions/.
[29] Dynamic Crowd Measurement. *DCM Features*. URL: https://web.archive.org/web/20240518035603/https://www.dynamiccrowdmeasurement.com/features.
[30] Andrew Tatrai and Travis Lachlan Semmens. *REAL-TIME CROWD MEASUREMENT AND MANAGEMENT SYSTEMS AND METHODS THEREOF*. Apr. 2022. URL: https://patentscope.wipo.int/search/en/detail.jsf?docId=GB357582318&_fid=WO2021011992.
[31] Datatilsynet. *Vejledning om dataansvarlige og databehandlere*. URL: https://web.archive.org/web/20250126130140/https://www.datatilsynet.dk/Media/7/6/Dataansvarlige%20og%20databehandlere.pdf.
[32] European Data Protection Board. *Guidelines 07/2020 on the concepts of controller and processor in the GDPR*. URL: https://web.archive.org/web/20250429030948/https://www.edpb.europa.eu/system/files_en?file=2023-10/EDPB_guidelines_202007_controllerprocessor_final_en.pdf.
[33] Intersoft Consulting. *Art. 28 GDPR - Processor*. URL: https://web.archive.org/web/20250421040340/https://gdpr-info.eu/art-28-gdpr/.
[34] European Commission. *European Artificial Intelligence Act comes into force*. URL: https://web.archive.org/web/20240910070941/https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_24_4123/IP_24_4123_EN.pdf.
[35] EU Artificial Intelligence Act. *Section 2: Requirements for High-Risk AI Systems*. URL: https://artificialintelligenceact.eu/section/3-2/.
[36] Microsoft Azure. *Pricing Calculator*. URL: https://azure.microsoft.com/en-us/pricing/calculator/.
[37] Qi Wang et al. "NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.6 (2021), pp. 2141–2149. DOI: 10.1109/TPAMI.2020.3013269.
[38] Yingying Zhang et al. "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 589–597. DOI: 10.1109/CVPR.2016.70.

[39] Vishwanath Sindagi, Rajeev Yasarla, and Vishal Patel. “JHU-CROWD++: Large-Scale Crowd Counting Dataset and A Benchmark Method”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP (Nov. 2020), pp. 1–1. DOI: 10.1109/TPAMI.2020.3035969.

[40] Shaokai Wu and Fengyu Yang. “Boosting Detection in Crowd Analysis via Underutilized Output Features”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2023, pp. 15609–15618.

[41] Joseph Redmon et al. “You Only Look Once: Unified, Real-Time Object Detection”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

[42] Ultralytics. *Explore Ultralytics YOLOv8*. URL: https://web.archive.org/web/20250426153402/https://docs.ultralytics.com/models/yolov8/.

[43] Tsung-Yi Lin et al. “Microsoft COCO: Common Objects in Context”. In: *Computer Vision – ECCV 2014*. Ed. by David Fleet et al. Cham: Springer International Publishing, 2014, pp. 740–755. ISBN: 978-3-319-10602-1.

[44] Maxim Tkachenko et al. *Label Studio: Data labeling software*. Open source software available from https://github.com/HumanSignal/label-studio. 2020-2025. URL: https://github.com/HumanSignal/label-studio.

[45] Yifu Zhang et al. “ByteTrack: Multi-object Tracking by Associating Every Detection Box”. In: *Computer Vision – ECCV 2022*. Springer Nature Switzerland, 2022, pp. 1–21. ISBN: 978-3-031-20047-2.

[46] A. Torralba, P. Isola, and W.T. Freeman. *Foundations of Computer Vision*. Adaptive Computation and Machine Learning series. MIT Press, 2024. Chap. 41. ISBN: 9780262378666. URL: https://mitpress.mit.edu/9780262048972/foundations-of-computer-vision/.

[47] OpenCV. *Camera Calibration and 3D Reconstruction*. URL: https://web.archive.org/web/20250327170811/https://docs.opencv.org/4.x/d9/d0c/group__calib3d.html#ga4abc2ece9fab9398f2e560d53c8c9780.

A Appendix

A.1 Workshop scenarios evaluation

	Scenario	Perceived value (1-5)	Effort reduction (1-5)	Confidence enhancement (1-5)
	1	4	5	3
Mads	2	4	5	4
	3	4	3	4
	1	4	5	4
Niels	2	4	5	4
	3	3	3	4

Table A.1: Evaluation of the workshop scenarios. The perceived value, effort reduction, and confidence enhancement are rated on a scale from 1 to 5. The ratings are based on the feedback from Mads Therkildsen and Niels Laustsen from Roskilde Festival's safety team.

A.2 Roskilde Festival meeting notes

See relevant meeting notes from the Roskilde Festival safety team below. Potential breaches of confidentiality and irrelevant information have been redacted.

A.2.1 February 16, 2024

```
Roskilde currently uses GPS phone signal to get overview of guest counts
through their app. Updates every 15 minutes to get a signal.

Morten says there are many systems available, but all come with a disadvantage
 - in Denmark, they dont use many cameras, as they are expensive to set up. On
 top of that they need to move them and this XXX up the analysis done by
cameras.

**Morten tested XXX**, they use a grid-like approach to calibration of cameras
. It can tolerate some camera movement but not much.

- Gratest value of XXX is in his eyes is the after-analysis of events. For
example to see how their barrier design influences flow.
- How fast can a scene get emptied, what is the flow of people from A to B.
How fast can a scene fill up and what are the connections to the temporal
aspects of the show (concert starting etc.)
- NOT interested in real time analysis - not even in front of stages, as there
 are plenty of people there present as ground crew and they can tell.
- Sometimes, the crowd issues happen outside of view of cameras (example of
Ukendt kunstner concert at Arena stage) far away from stage

It is crucial to build this product/app/function in collaboration with
educated crowd management professionals, and the accuracy of reporting is key.
 Following the current industry standards of flow estimation, density
calculation in pre-planning of venues. Otherwise he will NOT use it.
```

Morten says, psychology of people has changed after corona - They go much more to concerts and they are more hesitant to stand close together.

Festivals are scaling operations - during festivals they have a lot of employees, after festival everybody leaves and with that also knowledge of what was done last year. Creating documentation of layouts, crowd management plans and general analysis of how things were will be of great value.

The Police does security - preventing harm where there could be intent of harm
Roskilde does safety - Accomodating people so there is no harm in first place

Safety is about creating risk assessments for possible situations.

Communication: two separate channels 112 for emergencies and 114 for internal communication. They use VHF and Whatsapp groups. Sometimes phone signal is an issue.

Mads Therkilsen is responsible for flow analysis, and early in researching how AI tools can help with crowd management. (feeding existing information about venues to the …model.)

We will recieve contacts to people from england doing Crowd Science. Keith Still - mathematician with focus in this field.

Takeaway is - analysis is more important than real time insights

RF meeting notes – February 16, 2024

A.2.2 March 10, 2024

Crowd management and crowd safety is something some companies do, and some don't. In any case, to get licensed to make an event, you need to follow standard guides for a risk/event safety assesment that is submitted to the police. Issue is, the standard guide is very high-level and 'doesnt go in depth of what the actual safety plan is. Therefore, things are up for interpretation.

XXX 'doesnt plan, and they probably 'dont have the capacity or knowledge to do it either - their safety management is security focused, hiring people to manage things on the go on the ground as they develop. Sometimes, Roskilde offers some small planning if they are participating, however only internally.

Roskilde goes one step further by developing a Venue Manual for each stage/area of the festival. What they are interested in is understanding where people flow in from and mapping risks caused by these flows. They want to have a good understanding of what the flow of people over the whole venue is. People at every festival are different - at Roskilde they are younger, active, move around a lot and see many concerts.

UK is on the forefront of crowd management because they are a XXX up country with many problems, so they have to invest into it.

This summer, Roskilde only provides manpower on ground for XXX, no planning. Us making a plan for XXX / XXX since they don't have people that can do it?

Roskilde makes a **risk assessment for every single concert** - Starting with talking to bookers to understand what kind of a concert it will be (qualitatively speaking). After that, they research the band and do a **Band analysis** (what is the level of popularity/publicum attraction, what sound level they play at, what is the demographic). They try to see if they need to prepare for particular **crowd dynamics** (if the band causes moshpits or storming of the stages to happen etc.) **Crowd types are outlined by Berlongi ** [-Crowd types (Berlonghi 1995/EMA 1999) | Download Table (researchgate.net)](https://www.researchgate.net/figure/Crowd-types-Berlonghi-1995-EMA-1999 _tbl2_224911893)

They are building a database of bands and concerts to refer to. Word files with filled out features of the band/concert. Can we digitalize/log this data in our product?

Work flow:

Band Analysis > Risk Assesment > Concert Colour (red, green, yellow)

DJ sets are the most unpredictable.

Density mapping is very interesting. Where people leave and at what time, what direction. Setting up more cameras is only a problem when they are CCTV due to cost. Regular cameras is not a problem.

As planners, they have to decide based on their knowledge and intuition what densities are acceptable. There are EU guidelines, but in the end high densities are okay as long as people are having fun at a good concert.

They look at live camera footage to see if there is a crowd collapse, if people are …happy This is included in post evaluation.

Morten spends time looking at CCTV footage even after festivals.

What would you change if you had the product/analysis - Calculating number of emergency exits, which areas to close off and at what time. Staff allocation - better placement of people on ground in preparation of influx of crowds. Proving their theoretical knowledge with real data.

They follow the Event Safety Guide and the purple guide [The Purple Guide](https://www.thepurpleguide.co.uk/index.php/the-purple-guide)

[The+Event+Safety+Guide.pdf](Roskilde%20Festival%20 c0150926974744a3a5d39167c77340a2/TheEventSafetyGuide.pdf)

For planning of Roskilde festival, they use GIS map tool (Its also used by the emergency response departments). They can import AutoCAD drawings. Its a map of all services in one place. When they go out and build on site, they pay a land surveyor to map out the whole venue with accurate GPS coordinates.

RF meeting notes – March 10, 2024

A.2.3 May 25, 2024

Meeting with Mads Therkildsen.

Roskilde proposes our involvement at the festival to be composed of two phases . Pre-festival during warmup days they would like us to analyze EOS. After opening of the big festival, Arena stage. We are open and flexible to analyze also other parts of the venue, maybe areas with Food and Beverages.

Mads asks to be sent the project description, and to understand better what our aims with the product are. Ideally in the future, we are coming, setting up cameras, taking them down and just giving a report on the analyzed areas, with minimum work effort needed from the organizers side.

The expectation is to get data/demand curves for the ingress ang egress of people to the individual stages. Secondarily, to get relatively accurate measurements into what the density of crowds is in the analyzed zones that follow Fruins levels of service. Estimate density based in people pr. m2 etc.

It is interesting for Roskilde to get an idea about how a concert starting at Orange stage affects people moving from Arena to other areas.

Our contact people at Roskilde for the beginning will be Niels - working with CCTV systems and Adam from England who is a flow manager. We will setup a way to follow and learn from Adam while on ground working to understand the way they handle things.

We will receive risk assessments and concert schedules to understand key events.

RF meeting notes – May 25, 2024

A.2.4 August 28, 2024

Agreed on data exchange - week between 16th and 20th of September. Follow up meeting regarding progress to be agreed in October.

Soft milestones :

- 14th October - WIP Mockup for ingress/egress
- 25th November - Ingress/egress visualization in user-friendly platform and WIP mockup of crowd flow visualizations
- 20th January - Flow visualization in user-friendly platform integrated

RF is interested in seeing visualizations of direction of flow. Crossflow and density is bad for sales and influences their decisions when planning the placement of bars etc. We will ask for contact to a BI officer at RF to understand more how this data can be used in other wars than only crowd safety planning.

Key questions crowd officers ask: Do we have enough time to move people from one area to another? Low flow means higher densities. Direction of flow is important because they want to know where do people come from.

While they can use time to look at CCTV videos, it is more valuable for them to have a user friendly data based interface that can be understood across the entire organization. They could use the data in a broader sense to place bars , information screens, update their risk assessments, booker assessments of the attraction of the individual concerts.

RF meeting notes – August 28, 2024

A.2.5 September 18, 2024

Agenda:

Showcase the dashboard (cumulative total chart, ingress/egress chart)

Notes:

Nice visualization, only feedback on colour scheme

Morten can use the trends but not the absolute numbers, he wants to know what the error rate is, ho we validate it.

Bookers would LOVE this data

Morten wants a some highlights of the data that he can then use for making changes to the layout and staffing etc.

Overlay concerts from Gaia over this graph too

RF meeting notes – September 18, 2024

A.2.6 February 17, 2025

Levels of service

- Density measurement
- Heads and shoulders -> high density
- Feet, low density -> ~1-2 ppl pr m2

Flow

- Mads knows he needs 2500 guests in before 18
- Mads can count at 16:00, the ingress flow, and capacity of the entrance
- Multiply by an hour. How many come in an hour
- Also 15-min interval
- If Only 1000 people/hr, Need to increase capacity somehow
- 1-minute, 15-minute, 1-hour intervals most interesting

Mads sees a value in having continuous values

He only has initial count, not after changing capacity/flow

For planning purposes at concert:

- Mads needs to check previous flow numbers of venue to know when he needs to up/downscale flow at the entrance
- At Arena, constantly adjusting flow, redirecting crowds
- This data can help them see that 30-min before that need to redirect people Or here there is a density that is too high

Resource allocation is a huge part of it. 'Its hard to predict. They use estimates and experience to guess

When to open line up system at scene so that not too many come all at once
Also used to calculate entrance width. Based on experience.

Mads agrees that this tool helps with communicating and knowledge sharing. Internally across departments it helps justify requirements.

Map

- Mads likes the detail-level
- They have had heatmaps before. From app data. Every 15 minutes
- Only overview, traditional heatmap. Only colors/blobs. They can see that there are a lot people based on blob color
- Makes sense around food and beverage. Can see line formation

Density

- Mostly discussed in crowds in front of scenes
- Mads prioritises areas they know are high-density, to do their density calculations.

```
- 1.5-2 people pr m2 ish for food & beverage areas
- They want to know if there are areas with high density, where not designed
for high density
- Density numbers are still good. Shows real world situation vs. Planned
calculations
- You just 'dont want to see a high density.
- The data is documentation

**Other**

- Mads would like to see movement patterns.
- More specifically: where are people coming from, and where exactly are they
going
- Important to have people hit scene head-on, to prevent imbalance
```

RF meeting notes – February 17, 2025

A.2.7 March 24, 2025

```
**Flow**

- Mads likes the overview, 'its much more intuitive. He feels the numbers are
more useful
- They estimate flow to calculate how many meters are needed for proper
ingress/egress
- This helps document that space was properly allocated, or if it was too
narrow
- Would like to see flow separated for each entrance/camera
- Based on his intuition, the numbers look very realistic

- Would like to know the load of each entrance
- Flow chart should be able to filter based on entrance/camera, but also show
them aggregated

**Density**

- Chosen bins for heatmap are alright, but they use Levels of Service
internally, Mads sends a document

- This tool helps communicated internally in other departments
- Other departments dont understand crowd 'safetys recommendations
```

RF meeting notes – March 24, 2025

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Technical
University of
Denmark

Diplomvej, Buildings 371 & 372, 2nd floor
2800 Kgs. Lyngby
Tlf. 4525 1700

www.entrepreneurship.dtu.dk/