Automatic coordination of construction actors

Final report

Authors:

Olli Seppänen, Christopher Görsch, Jaakko Tuomisto, Joni Pajarinen, Yuan Zheng, Inshu Chauhan and Alaa Al Barazi (Aalto University)

Marja Liinasuo, Mika Hakkarainen, Kari Rainio, and Pekka Siltanen (VTT)

Contact in Aalto University: First Last first.last@aalto.fi

Contact in VTT: First Last first.last@vtt.fi

10.7.2024

Funded by the European Union – NextGenerationEU

Contents

1 Introduction

1.1 Background

ACTOR's premise is that a real-time, **digital situation picture** is the key to solving many construction problems. A trustworthy situation picture is a prerequisite for effective planning, task management, and logistics. When accurate situation data is stored in digital systems, work can be automatically coordinated, for example, by automatically assigning tasks to workers. A digital situation picture also enables each stakeholder in the project to have real-time visibility of all tasks, conditions, and resources.

The availability of data is no longer a challenge on construction sites. The digital footprint of construction has increased during the last decade. Cloud-based software applications, 360 cameras, fixed cameras, drones, and sensors provide real-time information about what is going on. However, a substantial portion of data management and analysis is manual, making providing a comprehensive situation picture difficult or impossible. The data are also contained in different systems in heterogeneous formats without a common language. It is very time-consuming to investigate all the different systems and try to comprehend the situation from digital sources.

The data are also subject to human interpretation and manual data entry. The current construction sites have a large amount of inaccurate data. The possibility of making errors should be decreased in all parts of the process. Data should be available automatically without manual entry, inferred from transactions that can be automatically sensed based on digital data streams.

Therefore, ACTOR envisions a future where the obtaining and distributing of situation information is automatic. Every project actor, from designers to field workers, has the correct information at the right time at their disposal in an easy-to-use visual form. This requires advancements in robotics, sensors, AI, machine learning, extended reality, and standardized data.

1.2 Work packages of ACTOR

To achieve the aim, the work in ACTOR has been divided into six technical work packages and a business model work package. Figure 1 illustrates the work packages and their relationships. This document focuses on the results of work packages A-F, which will be briefly introduced below.

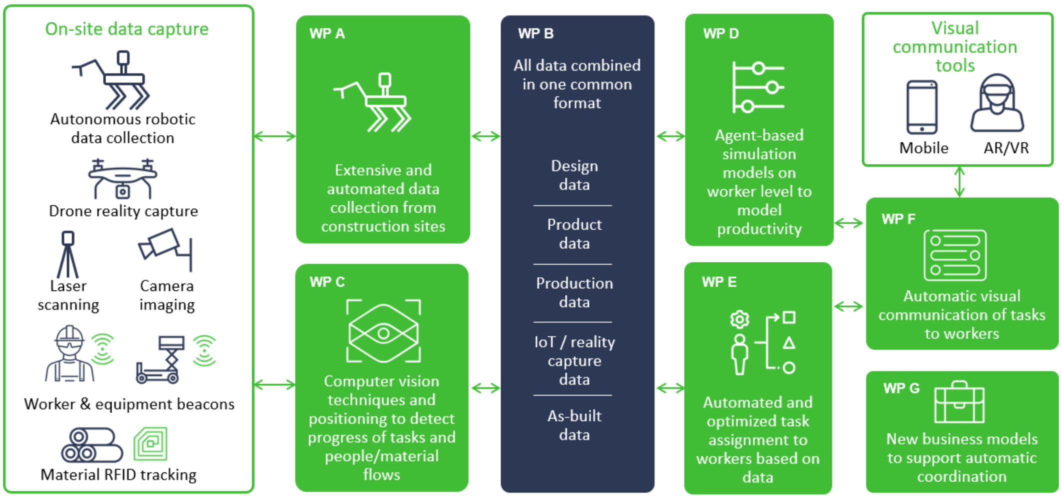


Figure 1: Work packages of ACTOR and their relationships

Automated data collection from construction sites (Work package A)

In Work Package A, digital data was collected from four construction sites. The collected data included resource positioning and condition data (with sensors), material delivery information, quality issues, schedules, laser scans, 360 videos and photos. The data were used as the starting point for other work packages. The work package contributes information about which data are important and which data collection activities can be automated.

The major focus of the work package was automated data collection. The research focused on solving practical obstacles related to using the Boston Dynamics Spot robot on site. Decision-making algorithms were developed to enable Spot to autonomously plan data collection routes. The goal is to have a robotic data collector who goes automatically without human intervention to collect visual data (images and point clouds) of areas that are under construction.

Standardized data representation (Work package B)

In this work package, the Digital Construction Ontologies (DiCon) (Zheng, Törmä & Seppänen, 2021) were expanded to accommodate new use cases required by this project and to enable standardized data representation for AI development. The work focused on storing image data in a consistent way, on rule-based reasoning about complex structures, and how to enable agent-based simulation with an ontology-based approach. As a practical output of this work package, all data collected in WP A were stored in a data lake, which conforms to the expanded DiCon ontology. This made all the collected data available for other work packages.

Analysis of situation data in images with computer vision (Work package C)

The work package focused on the progress detection of drywall activities using computer vision. Drywall was selected as the case because all the stages of drywall installation are not included in the Building Information Models, so it is not enough to compare BIM models to point clouds with occupancy-based methods. In addition to detecting the progress of each wall, we aimed to develop algorithms to find and label drywall-related materials and equipment from site images.

Digital twin of processes (Work Package D)

An agent-based simulation model was created for the drywall installation process. To model worker behavior, previously collected helmet camera data captured from the worker's point of view, as well as worker interviews, were used as the starting data. Experiments were done to achieve a digital twin by linking the simulation model to indoor positioning data. In this project, a proof-of-concept system was developed. In future research, the system will be validated by using it on real construction projects.

Decision optimization (Work package E)

The goal of this work package was to have a task list for each worker and continuously prioritize it so that selected tasks will only contain tasks that can really be done based on digital data sources. A worker decision-making model was defined based on factors that workers use to decide on the next task in practice. Task selection is critical to visually communicating the tasks to workers in WP F.

Digital visual management (Work Package F)

In the last technical work package, the decisions optimized in Work Package E were communicated to the worker with all required supporting knowledge, such as drawings, material storage areas, work instructions, etc., as visually as possible. Different technologies were tested for visual management and validated with construction workers on construction sites.

2 Overall ACTOR solution

ACTOR aims to enhance the management of daily operations on construction projects by improving the digital situation picture on site, ultimately enabling workers to proactively adapt to site dynamics and choose among available work options. ACTOR consists of four main modules, as presented in Figure 2. The first module is the Sensing System which aims to capture the actual situation on site continuously. This module relies on several technologies to generate a situation picture: mainly Bluetooth tracking beacons, 360 cameras, computer vision, laser scanning and point clouds, and other condition sensors like humidity and temperature sensors. The targeted elements for the situation picture are the elements under construction (e.g. phases of drywall installation), the material used for construction, the labor doing the work, and the site environment affecting the process.

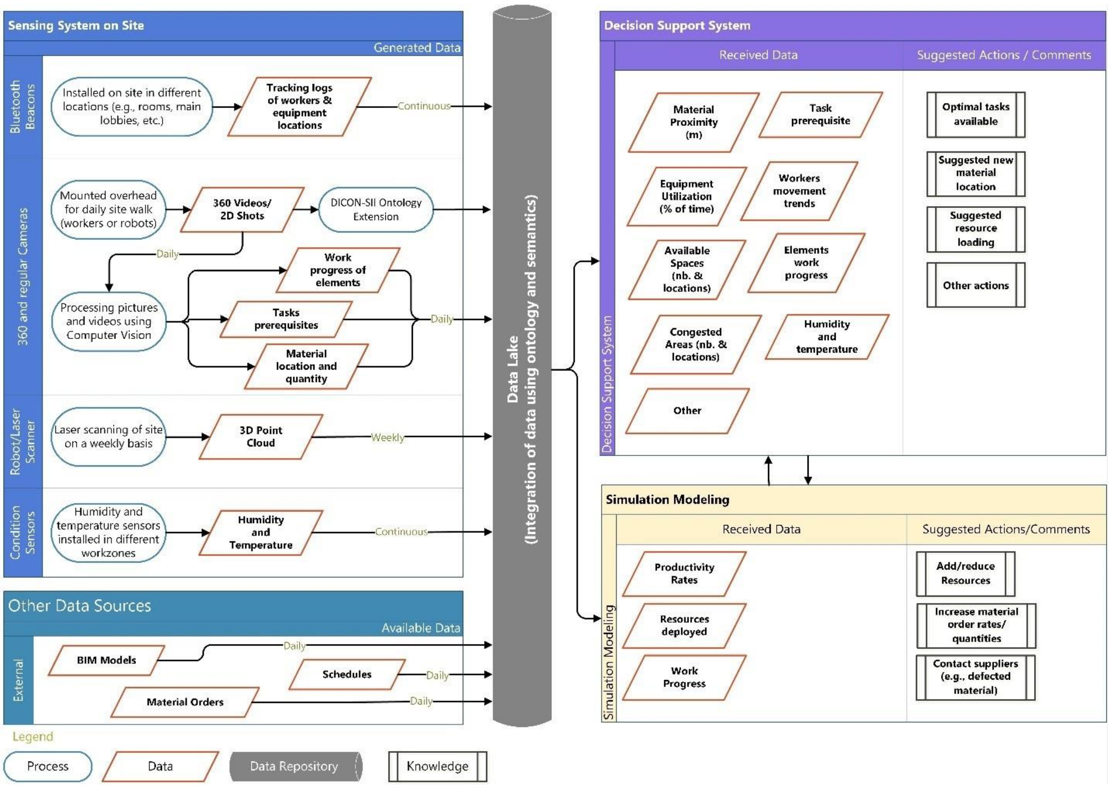


Figure 2: ACTOR Main Modules

The second module is the Data Lake, an ontology-coded database that collects data from the sensing system and other data sources, including BIM models, schedules, and material orders. Ontology is employed to allow for automatic queries from the database where different input information can be connected logically. The data lake is used to feed the simulation model and the decision support system. For example, by combining data from computer vision that tracks construction progress (e.g., drywall construction steps) with the data coming from beacons that track workers' location, the production rate of workers can be estimated for each step of the corresponding construction activity (e.g., framing, first layer, electrical, second layer, plastering, painting). The simulation model can use the value of the current production rate to estimate the time needed to finish the remaining drywalls. If delays are anticipated, managers can take action proactively to cope with schedule deadlines.

The third module is the Simulation Model, which mirrors the actual site conditions based on the input received from the sensing system through the data lake. The simulation model is built based on the layout geometry of the shared BIM model and updates continuously based on received data from the site and other data sources (e.g., client design changes). The simulation is used to predict the project's future performance based on the site's current status, including the actual number of workers, their production rates, material availability and storing locations, space usage and congestion, and workflow interruptions. In addition, the simulation can be used to test actions before implementing them. For example, simulation can be used to check the effects of changing the material storage location on movement waste and work productivity before deciding on the new storage location.

The fourth module in ACTOR is the Decision Support System (DSS), which aims to help workers decide on their daily activities based on actual site conditions. DSS aims to make workers more autonomous in managing their daily activities while sparing more time for superintendents and foremen to plan, check, and control tasks. Workers can have several work options to choose from every day. However, some activities may be better to optimize productivity, reduce movement, or enable inter-trade flow. The DSS will help workers select the most beneficial activities that make sense for the project and the worker's preferences. For instance, if two activities are available on a floor, the DSS will recommend starting with the one closer to the material location to reduce movement. Another example can be selecting an activity in a less congested area wherever several teams are working.

3 Description of collected data sets

Throughout the project, digital data was collected from projects with various technologies. The data were used to develop and validate the ACTOR solution. The researchers collected 360 video material and point clouds when testing the Spot robot. Other data sources came from the General Contractor of the project and were limited by the technology available in the project.

3.1 Digital data collection

Table 1 digital data collected from projects for use in work packages A-E. The data were collected with the General Contractor's consent on a campus building within Aalto University and three multi-family residential projects in the Helsinki region.

Table 1: Digital data collected during the project

Project name	Project type	Time period for collection	Collected data sets
K3 project	Campus	07-2022 to 10-2022	• BIM model in IFC format • Floor plan • Schedule • Images • 360 videos • Site arrangement • Quality control information • Cloud point model
Opaalikuja	Residential	03-2023 to 07-2023	• BIM model in IFC format • Floor plan • Schedule in SiteDrive • Images • Material list • Cloud point model
Postipuisto	Residential	09-2023 to 11-2023	• BIM model in IFC format • Floor plan • Schedule in SiteDrive • Images • Indoor position data for furniture installation • Cloud point model
Retiisikuja	Residential	11-2023 to 01-2024	• BIM model in IFC format • Floor plan • 360 videos • Schedule in SiteDrive

Data collection on the K3 project at the Aalto campus kicked off the project. Static data was collected mainly from there, including the BIM model, floor plan, and schedule. Researchers collected 360 videos and images. Remote-controlled laser scanning with a robot was also tested in the project. The aim of the next project, "Opaalikuja," was to collect also indoor positioning data. However, this failed due to a lack of consenting drywall workers. Thus, the data types collected from the Opaalikuja project were similar to the K3 project. "Postipuisto" project was another multi-family residential building with similar datasets. In this project, furniture installers consented to indoor positioning, and indoor positioning data were collected for the trade for two weeks on three floors. Finally, static data was collected on another residential project “Retiisikuja”.

4 Requirements

Each work package started by reviewing requirements for the scope of that work package. Requirements were identified through collaborative workshops with project participants. Some requirements resulted from learning from data collection and system development. The final understanding of the requirements gained during the project is described below, and it is related to digital situation picture, connectivity, and automatic data collection.

4.1 Required digital situation picture improvements

To achieve the aim of automatic coordination, digital situation picture needs to be created automatically based on the data contained in the data lake. Automatic coordination requires the real-time knowledge of work that is available for workers, as well as all the needed information to do the work. The agent-based simulation also requires more detailed knowledge of the process, including how fast the work progresses and how much effort is wasted.

Construction schedules are often not good sources for real-time process information. Although modern applications allow entering progress data in real-time with mobile devices, there are problems with the quality of data (Zhao et al., 2021). Workers do not consider their job to be filling in data in applications, so they forget or enter data afterward. The data can be partially correct related to current progress but cannot be trusted for productivity calculations. The situation assessed based on data in apps is often too pessimistic (workers have failed to enter progress that has been done) but often also too optimistic because workers have not understood the definition of done. Using self-entered information as the sole source for automatic coordination would result in several failures and dissatisfaction with the system. However, the schedule is important context information that can help interpret data contained in other systems.

To get more detailed process information, earlier research has attempted to use location tracking of workers to evaluate the progress of tasks. The overall approach is to observe when workers leave a location and move to the next location and determine start and finish times based on worker presence. Dependencies and schedule information help to interpret the data and to assess the tasks being completed correctly. If only longer periods of uninterrupted presence are considered, this method gives more accurate start and finish times than manual entry in systems (Zhao et al., 2021). However, without combining the information with other systems, it is difficult to evaluate accurately what has been done or to understand the detected patterns of movement.

The only objective way to understand what has been completed is to inspect the location visually. Because ACTOR attempts to minimize the amount of human labor involved in automatic coordination, the requirement is to do the inspections automatically. Automatic inspections can be done with robots. However, directing robots or creating data collection plans should also be avoided. The robot should automatically go to collect data in areas where progress has been made based on other data sources. The key data sources include indoor positioning systems, BIM models, and schedules. Workers leaving the work area and moving to work somewhere else should automatically alert the robot to collect visual data from the area. Data should be collected related to all BIM elements that pertain to the task in question to get element-level progress. This leads to the following requirements:

1. Indoor positioning should tentatively detect the start and finish times of tasks (e.g. based on Zhao et al. 2021) based on the construction schedule

2. BIM model elements should be linked to construction tasks (I.e. a 4D model should be in use)
3. The robot should start a visual data collection mission when uninterrupted presence is detected in a work location, or workers move to the next work location
4. Data collection should be done related to all elements related to the task in question

If these requirements are fulfilled, visual data, such as point clouds and images, are collected related to each element of interest. This data should then be automatically analyzed to detect progress on the element level. For some elements, comparing point clouds to BIM models with occupancy-based methods (e.g., ductwork, cable trays) is enough. Other elements are typically not included in the model on such a level of detail that work stages can be seen. For example, drywall includes several stages 1. layout, 2. framing, 3. hanging one side of drywall, 4. electrical and plumbing work in walls, 5. insulation 6. closing the drywall, 7. taping and finishing and 8. painting. Typically, drywall is included in Level-of-Detail 300 BIM models with just one element. For example, the studs are not included in a typical architectural model. However, scheduled tasks in the schedule are often on a more granular level, and different stages are done by different crews (e.g., steps 1-3,5, and 6 by carpenters, 4 by electricians and plumbers, 7 by plasterers, and 8 by painters).

To accurately detect stages of drywall, point clouds can be used as a starting point. 2D maps can be generated by the robot to see if something has been installed in the location of the wall. The actual stage should be evaluated by appearance-based computer vision techniques, such as deep learning. The requirements coming from this are as follows:

5. Point clouds and occupancy-based methods should be used to detect the progress of simple elements that are on the correct level of detail in BIM models
6. Images and appearance-based computer vision methods should be used to detect the progress of complex assemblies that are not modeled in detail and work tasks that change appearance but do not create sizable new elements (e.g. floor coverings, painting)
7. Robot should continuously map the construction site to help detect areas where the work is done. A map should be created at least once daily and updated during any data collection trip.

Continuous robotic mapping also helps detect material storage areas and temporary installations, which are important parts of situational awareness. The digital situation picture system should keep track of temporary elements and analyze them using computer vision.

8. The maps created by robots should show temporary elements that take space
9. The robot should stop and collect visual information about any new temporary elements taking space that are encountered during data collection trips.
10. Computer vision should be used to classify and understand temporary elements, e.g., to recognize material storage areas, including materials or equipment, for a certain task.

In addition to detecting physical progress and materials, the system should detect whether conditions are good for continuing or starting work. This requires IoT sensors for each location and knowledge of the condition requirements of each task.

11. Condition requirements (e.g. temperature, humidity) should be known for each task in the schedule

12. Condition sensors should be reviewed to see tasks that can be done based on real-time condition
13. Alarms should be given to management when an upcoming task does not have the right conditions

Finally, the process itself should be analyzed on a micro level. Indoor positioning data can be used as a starting point for this analysis. In addition to spending time in work areas, the workers are moving to other areas, often for short visits (Zhao et al. 2019). Evaluating the frequency, duration, and reason for these interruptions would enable productivity improvements. Frequency and duration by destination can be analyzed by just looking at the data collected by the indoor positioning system. Understanding the reason requires reasoning based on other data sources.

14. Indoor positioning data should be continuously analyzed to find locations where workers visit during work. The frequency and duration of these visits should be measured.
15. The robot should go to collect data on any unknown locations (not including social areas for ethical reasons)
16. The locations should be classified using computer vision methods.

In ACTOR project, we have attempted to fulfill the requirements related to the drywall task. The overall process of forming the digital situational picture and how it connects to simulation models and visual management is depicted in Figure 3.

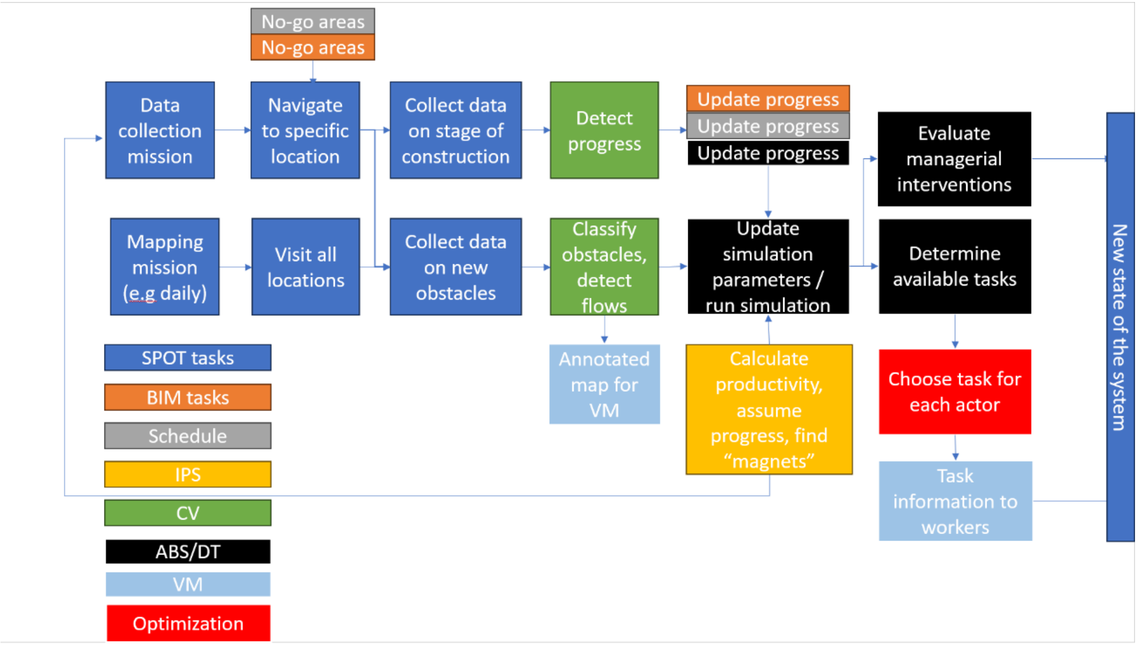


Figure 3. Process from data collection to visual management

Data collection mission

The indoor positioning system triggers data collection missions by Spot. It should collect data in work locations where uninterrupted presence by workers has been noticed for extended time periods (for example, more than two hours) since the last data collection. In addition, data should be collected in locations where workers visit regularly ("magnet locations"). There can be multiple destinations for any one data collection mission. For example, there could be several “magnet locations” to classify and several locations where workers' uninterrupted presence has been detected.

Mapping mission

Regularly (e.g. once a day), Spot will conduct a mapping mission of all project locations. In this mission, the aim is to update a real-time map of progress and obstacles on site. On the first visit, a BIM-based map is used as a base. On subsequent visits, the map is dynamically updated, considering only fixed elements that do not move during observation. Image data is collected if there is a new obstacle that has not been detected before or if new points are seen in locations where there are BIM elements.

Navigate to a specific location

Spot plans a route to the locations specified in the data collection mission. In route planning, Spot avoids any areas that have been labeled no-go areas in the data lake or where no-go activities are taking place. Any activity in the schedule can be labeled a no-go activity. No-go status can be active only when the work is ongoing or also for a period after the activity stops (e.g. to allow concrete to harden or floors to be covered). Spot also dynamically avoids people and obstacles on the route. If no route can be found to a destination, that part of the mission fails, but Spot attempts to complete other parts of the mission. Spot should collect data on any obstacles preventing data collection.

Visit all locations

SPOT plans a route that visits all work locations except for those labeled No-go (see above).

Collect data on the stage of construction

Data collection missions to work locations where uninterrupted presence has been detected aim to collect data on work that has been completed. Data on stage of construction can also be collected opportunistically on mission to visit all locations, if points are detected on BIM elements. For drywalls, this means capturing good-quality point clouds and pictures on all drywall elements. Data are collected for each BIM element separately and stored related to that element in the data lake.

Collect data on new obstacles

Data collection missions can identify fixed obstacles that are not in the BIM map. Pictures and point clouds should be collected from these obstacles.

Detect progress

Using data collected on the stage of construction, computer vision methods identify the stage of each element (drywall in the ACTOR project). The resulting stage is updated in the data lake.

Classify obstacles, detect flows

Spot-detected obstacles are analyzed with computer vision methods. Construction materials are detected and quantified. In the ACTOR project, detection is limited to drywall materials. Their information is stored in the data lake.**An annotated map for VM**

The map created for Spot is taken as a starting point and annotated with data contained in the data lake. Any magnet locations should show how often workers visit that magnet and how long they stay. The map should also provide more detailed pictures.

Update progress (in simulation)

Simulation uses data in the data lake to update floor plan obstacles and their locations. The initial assumptions for simulation are updated to match their real status. Progress detection supplies the stage of each drywall, allowing the model to calculate the amount of work left for each task for each element.

Update simulation parameters / run simulation

The simulation model calculates productivity by combining progress data with indoor positioning data. Progress data has a time stamp of data collection. The manhours of presence from the indoor positioning system for workers working in the stage are divided by the output in square meters to calculate the labor consumption related to each stage. This labor consumption is used to forecast the system's future behavior.

Calculate productivity, assume progress, find “magnet locations”

Before Spot has collected data, indoor positioning data is used to “assume progress” using the process of Zhao et al. (2021). Task switches in a location are determined by the patterns of beacon movements in the room to determine started and finished activities. This information is used by the simulation model until Spot has collected actual data on progress. Indoor positioning also identifies magnet locations, which are locations that are visited by workers regularly during their work.

Evaluate managerial interventions

Managerial interventions that can be evaluated with the system include:

- Additional / fewer resources for tasks
- Eliminate magnet locations (for example, by getting additional resources to carry materials close to the installation location) or change their location (moving storage area close to the installation location)
- Overtime work
- Changing the sequence of construction

Determine available tasks

Available tasks are those with materials available (detected by Spot and classified by CV), workers available (beacons assigned to tasks), and previous stages completed.

Choose a task for each actor

The best task is selected from available tasks for each actor corresponding to a beacon. Developing a task selection algorithm for construction processes, such as drywall installation, involves several factors. These factors include proximity, work amount, work left, historical congestion, and the involvement of other trades. A task selection algorithm can optimize efficiency and effectiveness by prioritizing tasks based on these factors.

Task information to workers.

Task information is presented visually to workers using AR and info screens.

4.2 Requirements for connectivity

Creating a digital situation picture requires access to large amounts of data that should be accessible to other systems. Because single data collection devices cannot easily combine information from various sources, internet connectivity with sufficient bandwidth and low latency is required. The primary purpose of the internet connection is to transmit autonomously captured data, which includes IoT data, images, videos, point clouds, and app data. The following requirements are based on experience of field tests in case projects.

4.2.1 Internet Connectivity on construction sites

It is unlikely that a wired internet connection will be available at the construction site during the construction phase. Thus, it may be necessary to use a wireless mobile broadband connection. Unlike fixed-line internet connections (e.g., DSL, cable, fiber), mobile broadband connections do not rely on physical infrastructure like wires or cables to transmit data, making them suitable for deployment in construction sites.

A modem is a required component of mobile broadband, but its placement can vary depending on the approach. There are two main approaches to modem placement: (1) a decentralized approach, where the modem is installed on a data collection device, and (2) a centralized approach, where a single modem is installed at the construction site and the data collection devices communicate with it.

The centralized modem approach offers the ability to support multiple devices or robots simultaneously. However, the centralized approach also requires a Wi-Fi network with good coverage at the construction site, as devices must be connected to the modem via Wi-Fi to transmit the collected data. Because of this, the decentralized approach enables greater independence for devices, as it is less dependent on the availability of infrastructure on the construction site (the mobile broadband router and Wi-Fi coverage).

Another critical aspect that needs to be considered is the placement of the mobile broadband antenna. In a centralized setup, the antenna can be directional, larger, and placed in an optimized position and orientation, providing a better signal strength and, in turn, potentially more reliable connectivity. In contrast, when the modem is attached to the data collection devices, the antenna moves if the device is mobile (such as a robot). This can lead to potential connectivity issues, as the cellular signal may have to penetrate multiple walls.

If a wired solution is available at the construction site, it may be a more suitable option due to the reliability and stability of wired connections. For example, elevators could be used as a hub for connectivity if they are installed early during the construction process. In this case, a centralized modem setup would be required to connect the robot to the wired internet via Wi-Fi.

4.2.2 Bandwidth Requirements

Several data collection devices, such as robots, require reliable connectivity to transmit the collected data. The most bandwidth-intensive part of the connectivity requirements are the images and point cloud data, which can be captured by sensors aboard robots, fixed or mobile cameras, or laser scanners. This is because images and point cloud data are typically much larger in volume than other types of data, such as sensor readings, application data, or robot control commands. Therefore, the estimates for bandwidth requirements will focus on the transmission of images and point clouds. Most manually collected data do not need to be uploaded in real time. However, data collected by robots in ACTOR is used to dynamically update data collection plans, so sensors on the Spot robot are considered in the following.

In the configuration used in ACTOR, point clouds are collected using a Light Detection and Ranging (LiDAR) sensor, the Ouster OS0 (64 beams variant). The LiDAR is used with a resolution mode that captures 512 points for each beam, making it a total of 32,768 points for each captured point cloud. The LiDAR captures point clouds with a frequency of 10 Hz and supplies data with a rate of 44.6 Mbps (~ 5.6 MB/s). However, it is not necessary to transmit each scan. For example, if we want to transmit one scan per second, we require a reliable bandwidth of only around 5 Mbps.

In addition to point clouds, images need to be transmitted. The robot has 13 different camera sensors, each of which captures images with a 10 or 15 Hz frame rate. To transmit a single image from each camera per second, we require a bandwidth of around 2 Mbps. In summary, we estimate that an internet connection with a bandwidth of at least 7 Mbps is required to transmit an update from each sensor from SPOT once per second.

4.3 Requirements for automatic and autonomous data collection

This section will outline the requirements for an autonomous data collection system that can capture images and point cloud data from a construction site. We will discuss the sensors and equipment needed to collect autonomous data. Additionally, we will explore the challenges and limitations of implementing an autonomous data collection system in a construction environment and outline the requirements for overcoming these challenges.

4.3.1 Sensor Requirements

Sensors are essential components of autonomous data collection systems. However, some sensors are more critical in the robot's ability to move and operate independently. For example, a LiDAR scanner can be considered essential for mapping and navigating the site, which is a requirement for the robot to move autonomously and accurately. Other sensors may be important for their data collection capabilities but are not so important to functional requirements. When selecting sensors for the system, these different priorities must be carefully considered.

For example, different types of LiDAR scanners have distinct characteristics that significantly impact their suitability for the application in terms of both functional performance and the quality of the collected data. 2D LiDAR scanners can capture simple depth information of the environment in a single plane. Their simplicity and relatively lower cost make them a popular choice for Simultaneous Localization and Mapping (SLAM) applications. While they can be effective in certain environments, one of the main drawbacks of 2D LiDARs is that they produce relatively uninteresting data in the context of construction sites, as they are limited to creating simple occupancy-grid-type maps. This can create challenges in both navigation and data collection. In contrast, 3D LiDARs can capture significantly more information, but they are also more expensive and more difficult to integrate into a robotic platform due to their larger size and weight.

Furthermore, not all sensors can be used with all robotic platforms. For example, the size and weight of a 3D LiDAR scanner make it unsuitable for deployment on many robotic platforms. Therefore, when selecting a robotic platform for autonomous data collection, it is essential to consider the functional requirements of the sensors and the capabilities of the platform. The Boston Dynamics Spot, for example, provides a robust platform for deploying a sizable LiDAR scanner, making it a viable choice for autonomous data collection.

To achieve the stated purpose of Work Package A, which is the autonomous collection of images and point cloud data, a LiDAR scanner and one or more cameras are necessary. The LiDAR sensor is essential for capturing the required 3D point cloud data, but it is also used to build a map of the environment and to localize the robot within that environment. The map is a critical component in achieving autonomous navigation, as it is used to plan paths that avoid obstacles for the purpose of autonomous data collection.

In this project, a Boston Dynamics Spot robot with its integrated sensors and an additional external LiDAR scanner were utilized to capture the construction site's images and point cloud data. Specifically, the Spot robot's sensors include:

- An RGB visible-light camera for capturing color images
- 5 monochromatic visible-light cameras for capturing grayscale images
- 6 depth cameras for capturing 3D depth information
- An infrared camera for capturing monochromatic images in dark conditions

Additionally, an external Ouster OS0 (64-beam) LiDAR scanner was installed on the Spot robot. This LiDAR was used to capture point cloud data of the construction site and to support autonomous navigation of the robot.

4.3.2 Capturing 3D Information

Capturing 3D information is an essential aspect of a real-time situational picture in construction environments. Different sensors, including cameras, depth cameras, and LiDAR sensors, can be employed to collect 3D data.

Normal cameras are not inherently capable of capturing 3D data. However, various techniques, such as structure from motion (SFM) or deep learning-based depth estimation, can be used to generate 3D data from the images. These techniques can provide a cost-effective and efficient way to generate 3D data, although depth estimation accuracy is typically lower than dedicated 3D capture methods.

Alternatively, depth cameras can be used, which rely on techniques like stereo vision (comparing images from two offset cameras), structured light (projecting known patterns and analyzing distortions), or time-of-flight (measuring round trip time of light pulses) to calculate depth. They allow for small and relatively low-cost implementations, especially in comparison to LiDAR sensors. However, while depth cameras produce dense measurements, their range tends to be limited (typically between 0.5-5 meters).

Finally, LiDAR sensors are typically the most expensive option, but they excel in terms of accuracy, range, and robustness to lighting conditions. They can physically measure distances up to hundreds of meters with high accuracy, a feat that is difficult or impossible for conventional depth cameras. For this reason, LiDARs are a popular choice for many applications, including SLAM. However, when compared with depth cameras, LiDARs are also substantially pricier and often unwieldy to integrate. LiDAR was selected as technology for ACTOR project because of the ability to get detailed 3D point clouds and the need to use SLAM for autonomous navigation.

When selecting a LiDAR sensor, several factors must be considered. These include the following:

- The LiDAR type (2D/3D)
- The number of beams or channels
- Horizontal and vertical field of view
- Measurement accuracy (angular sampling accuracy and distance measurement error)
- Measurement range
- Size and weight
- Power requirements

The choice between 2D and 3D LiDAR is arguably the most significant out of these factors. 2D LiDAR sensors capture depth data in a single plane. This can be considered insufficient for the monitoring of construction sites. While 2D LiDARs can create 2D occupancy grids, maps, or floor plans, this data cannot readily be used to distinguish between different obstacle types apart from simple scenarios. Moreover, 2D LiDARs typically capture fewer features in SLAM applications. This can lead to accuracy issues, especially in dynamic environments like construction sites where obstacles and walls can change location. In such cases, it can be difficult for a robot to navigate or understand the site accurately. In contrast, 3D LiDAR scanners can capture much more information, including features such as walls, ceilings, and other infrastructure. This makes them better suited for applications that require a more comprehensive understanding of the environment.

While 3D LiDAR scanners offer significant advantages in terms of capturing a more comprehensive understanding of the environment, they also have some drawbacks compared to 2D LiDAR scanners. One of the most significant limitations is their greater cost. 3D LiDAR scanners are typically more expensive than their 2D counterparts, which can be a significant consideration for projects with budget constraints.

Another issue of 3D LiDAR scanners is their increased size and weight, which can make them more difficult to integrate into smaller or more compact robot designs. Furthermore, the added weight and size can also impact the robot's overall performance, maneuverability, and runtime. Additionally, 3D LiDAR scanners require careful placement on the robot to ensure that they have a clear and unobstructed view of the environment. This can be more challenging than with 2D LiDAR scanners, which only need an unimpeded view in a single plane.

Given these considerations, 3D LiDAR is a requirement. However, there are significant differences even between 3D LiDARs that need to be assessed. An important consideration is the number of beams. The number of laser beams, also known as channels, plays a critical role in determining the density of the point cloud generated by a LiDAR sensor. A LiDAR with a single beam is essentially a 2D LiDAR, while an increased number of laser beams increases vertical resolution and yields a more densely packed 3D point cloud. Put simply, this suggests that we should aim for the highest number of beams or channels possible to attain the highest density and resolution in the generated point cloud data.

The field of view is another critical aspect to consider. While most LiDAR sensors have a horizontal field of view of 360 degrees, the vertical field of view can differ significantly depending on the intended application and measurement range. For example, LiDARs designed for long-range outdoor use typically have lower vertical fields of view, whereas those designed for indoor use typically have wider vertical fields of view. In the context of SLAM, a wide vertical field of view is particularly beneficial in indoor scenarios. A narrow vertical field of view can limit the sensor's ability to capture the entire space, making it challenging to localize the robot and maintain a robust understanding of the environment.

However, while a large vertical field of view is desirable, it typically comes with decreased measurement range and accuracy. In most buildings, autonomous data collection is mostly operated in close-quarters scenarios. This means that a lower measurement range does not have a significant effect if it is above a reasonable minimum (e.g., 30 meters). Conversely, the distance measurement and sampling accuracy of the LiDAR remain important and should be maximized.

Finally, it is essential to analyze the size, weight, and power requirements of the LiDAR. A larger and heavier LiDAR sensor may not be suitable for robots or vehicles with limited payload capacity or space, while a smaller and lighter sensor could be more practical. Moreover, the added weight and size of a LiDAR sensor can also affect the robot's overall performance, maneuverability, and runtime. Additionally, the power requirements of the sensor typically increase alongside other performance figures. Essentially, the power requirements of the LiDAR sensor should not exceed the available supply or significantly reduce the runtime.

For the purposes of Work Package A, the Ouster OS0 LiDAR sensor was selected. This specific variant of the Ouster OS0 features 64 beams and a 10 Hz sampling rate and was operated in the 512-point horizontal resolution mode per beam. Importantly, the sensor has a wide vertical field of view of 90 degrees, making it particularly useful for use in tight indoor spaces. Further, according to the manufacturer's specifications, the sensor has a range of 35 meters (10% Lambertian reflectance, >90% detection probability), angular sampling accuracy of ±0.01 degrees, and range accuracy of 5 cm. The sensor requires between 14-20 watts of power, with an operating voltage of 12 VDC, supplied from the Spot robot's battery.

In the used configuration, the Ouster OS0 LiDAR captures up to 32,768 points per scan at maximum, which occurs 10 times per second. This translates to a maximum rate of 327,680 points per second or 44.6 Mbps (~ 5.6 MB/s). In practice, the number of points captured is lower, often ranging between 15 and 30 thousand points.

4.3.3 Capturing Images

While LiDAR sensors are crucial for capturing detailed 3D data, cameras play a significant role in the ACTOR project for appearance-based detection of progress or materials.

For autonomous surveying of construction sites, 360-degree cameras and high-resolution industrial cameras are standout choices. 360-degree cameras, such as the Ricoh Theta or Insta360 Pro, capture panoramic views of the target site in a single shot without the need for multiple cameras or images. These cameras are typically relatively compact, lightweight, and mountable on robots. Conversely, high-resolution industrial cameras, such as those from Basler, FLIR, or Allied Vision, typically offer high-quality images with robust housing and connectivity.

While a high-resolution, high-quality camera system capable of capturing 360-degree images would be the ideal choice, we made a practical compromise to utilize the internal camera sensors of the Spot robot. These sensors provide a sufficient level of image quality to survey the construction site. Specifically, the used cameras and their features are as follows:

- An RGB visible-light camera for capturing color images
    - Installed on the arm of the Spot robot, looking straight ahead

  - Resolution of 640x480 pixels
  - Frame rate of 15 Hz
- 5 monochromatic visible-light cameras for capturing grayscale images
  - Installed on the body of the Spot robot, providing a 360-degree view around the robot (two at the front, one at the rear, one on each side of the robot). Looking downward, primarily used for observing the ground and low obstacles
  - Resolution of 640x480 pixels
  - Frame rate of 10 Hz
- 5 body depth cameras for capturing 3D depth information
  - Installed on the body of Spot robot, one depth camera in the same location and looking in the same direction for each monochromatic visible-light camera
  - Detection range ~2 meters
  - Resolution of 424x240 pixels
  - Frame rate of 10 Hz
- A depth camera on the arm for capturing 3D depth information
  - Installed on the arm of the Spot robot, looking straight ahead
  - Detection range ~2 meters
  - Resolution of 224x171 pixels
  - Frame rate of 15 Hz
- An infrared camera for capturing monochromatic images in dark conditions
  - Installed on the arm of the Spot robot, looking straight ahead
  - Has an associated infrared illuminator, providing images even in total darkness
  - Monochromatic
  - Limited range of around ~3 meters
  - Resolution of 224x171 pixels
  - Frame rate of 15 Hz

The cameras installed on the robot were used to capture visual images of the construction site, providing contextual information and augmenting the 3D point cloud data obtained from the LiDAR sensor while fulfilling the requirements set for data collection.

4.3.4 Additional requirements for data collection robots

As part of the research for Work Package A, we aimed to achieve a set of ideal features and capabilities that would enable a robot to operate effectively as an autonomous data collector on a construction site. The following requirements were identified:

- Safety
  1. The robot must operate safely and avoid collisions with people, equipment, and structures
  2. Fail-safe mechanisms should be in place to automatically stop the robot in case a problem is detected

3. The robot should have warning lights/sounds and a clearly visible emergency stop button

- Localization and Navigation
    1. The robot must be able to localize itself within the construction site accurately
    2. Autonomous navigation should safely avoid obstacles and reach all data collection locations
    3. The navigation system should be able to operate and navigate in multistory buildings and between floors
    4. The robot should be capable of navigating uneven terrain, moving over obstacles, and climbing/descending stairs and slopes.
    5. THE Robot should be able to operate in tight spaces as required
- Robot data collection features
    1. RGB cameras for collecting data
    2. LiDAR sensor for data collection and navigation
    3. Onboard computing resources with sufficient RAM, CPU cores, and SSD storage
    4. Wi-Fi or LTE/5G connectivity for real-time transmission of data
    5. Robust power management with a minimum runtime of 30 minutes and the ability to hot-swap or recharge in the field
- Off-site Infrastructure
    - Centralized data management system for storing, processing, and analyzing collected data
- Regulatory Compliance
    - Compliance with relevant construction site safety regulations and standards
    - Implementation of data security, privacy, and insurance requirements

5 Automatic data collection using Boston Dynamic Spot

The goal of ACTOR includes updating the digital situation picture with minimum human intervention. Some IoT systems, such as indoor positioning or condition measurement and fixed cameras, stay in fixed locations. However, it is not feasible to install fixed cameras in such a way that the whole construction site would be covered. Thus, some mobile data collection is required. Traditional methods, such as helmet-mounted 360 cameras, normal cameras or laser scanners, require a human operator. In the future, robots could be used for autonomous data collection. In this chapter, we introduce the robot we used for tests and how data collection plans can be automated.

5.1 Test robot setup

We selected the Boston Dynamic Spot quadruped robot for tests because it can navigate uneven terrain, move over small obstacles, and climb and descend stairs and slopes. For data collection, the robot is equipped with internal RGB and monochromatic cameras, which were used to capture images of the construction sites. Additionally, we installed an external LiDAR sensor (Ouster OS0) on a custom 3D-printed mount. The LiDAR sensor provided accurate distance measurements and point cloud data, enabling the robot to create accurate 3D models of the environment.

Two different Wi-Fi interfaces were used to facilitate reliable communication and data transmission. This includes the robot's internal Wi-Fi and an external Wi-Fi interface installed on the robot's back. The internal interface had limited bandwidth because it is restricted to 2.4 GHz Wi-Fi bands. The internal interface was used only for control commands. However, the external Wi-Fi module installed on the robot can operate on both 2.4 GHz and 5 GHz bands and achieve real-world data transfer rates of up to 600 Mbps. This allowed it to be used for real-time image and LiDAR data transmission. In terms of runtime, the Spot robot could operate for a full hour, exceeding the minimum requirement of 30 minutes. Figure 4 shows the testing setup of Spot.

In practice, having the required computational resources embedded directly within the robot would be ideal. However, due to the computational demands of the tasks at hand, we opted to utilize a powerful external computer to control the Spot robot. This computer was connected to the robot using Wi-Fi. The external computer processed complex tasks, including mapping, navigation, and deep-learning semantic segmentation. The used computer featured a Ryzen 9 7950X3D CPU, 128 GB of DDR5 RAM, and an Nvidia RTX 4090 GPU.

When operating in complex environments, safety is closely tied to the robot’s navigation ability. The most significant risk factor is the robot's ability to navigate and avoid obstacles, as this directly impacts its ability to operate safely. For obstacle avoidance in real-time, we utilize the Boston Dynamics Spot's internal obstacle avoidance and safety features. This internal obstacle avoidance system is designed to prevent collisions with people, equipment, and structures automatically. While this system is not infallible, it has proven to be effective in most situations, and the robot has not had any major collisions or accidents.

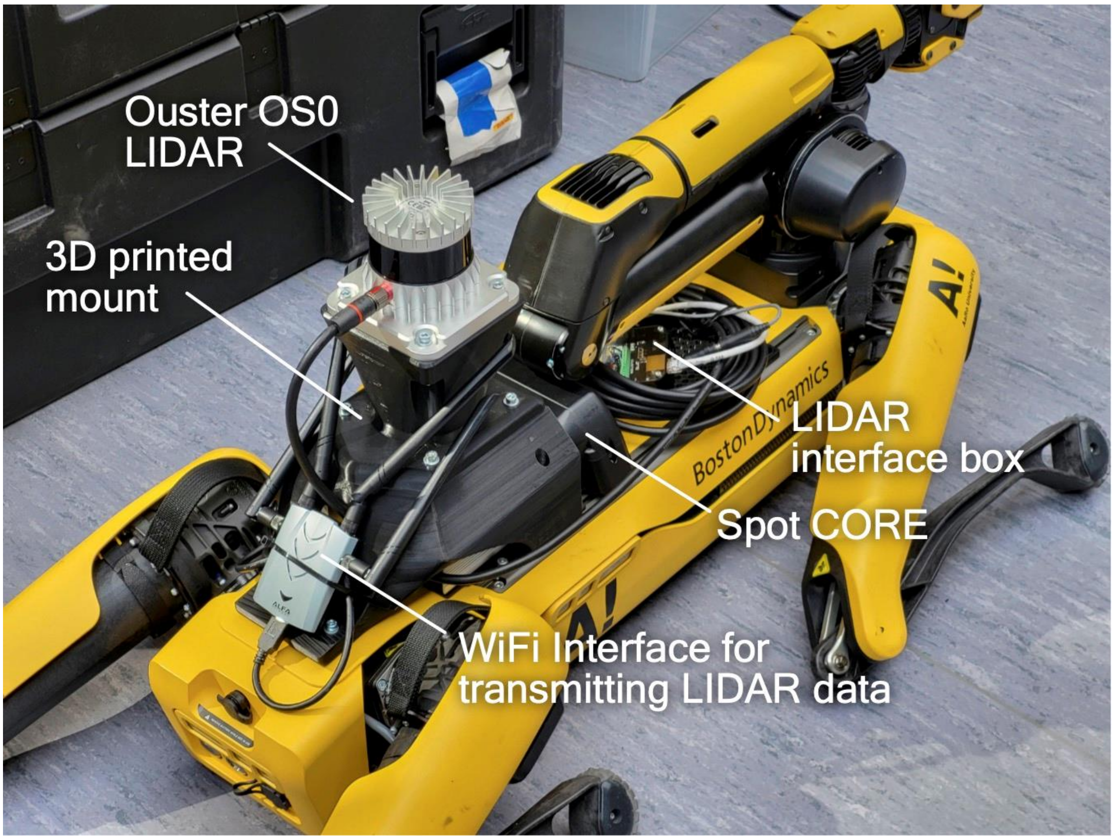


Figure 4: Boston Dynamics Spot, with an external LiDAR, Wi-Fi interface, and computer installed.

5.2 Localization and Navigation

To ensure accurate and efficient data collection, the autonomous data collection robot requires reliable localization and navigation capabilities. The robot must be able to pinpoint its location within the construction site, navigate to different data collection locations, and avoid obstacles and humans.

Navigation methods can be split into two categories: global and local planning. Global planning involves high-level planning for navigating from the starting point to the destination. Global planning methods generally leverage pre-existing knowledge about the environment, like building floor plans or a map generated by the robot's sensors, to find an optimal path. Local planning focuses on generating short-term, reactive behaviors to navigate the immediate surroundings, considering real-time obstacle avoidance and dynamic changes in the environment. Local planning methods rely heavily on sensor data to perceive and respond to changes in the local environment.

In practice, navigation systems often combine both global and local planning methods to achieve robust and efficient navigation. Our approach adopts a similar strategy and involves a combination of global and local planning in the navigation system. At the global level, the robot uses a metric map of the environment to plan its route and avoid major obstacles. This map is updated in real time, reflecting changes in the environment as they occur, such as new construction or changes to the layout of the construction site. In other words, the robot's global planner considers any known or anticipated obstacles and navigates around them. The map used by the global planner can be initialized with BIM data in the form of Industrial Foundation Classes (IFC) files to reduce or eliminate the need for mapping.

For navigation on the local level, we rely on the Spot robot's inbuilt trajectory following and obstacle avoidance features, which use Spot's internal sensors to detect and avoid obstacles in real-time.

5.2.1 Localization and Mapping

There is a fundamental connection between mapping and localization. Mapping relies on accurately pinpointing the robot's position and orientation within the environment, while localization relies on having a thorough understanding of the environment, which can only be achieved through the mapping process (SLAM).

We evaluated available open-source SLAM implementations to localize the robot, focusing on 2D and 3D SLAM. 3D LiDAR SLAM was the most reliable choice for our application. 2D SLAM was less reliable due to the lack of defining features in a single plane. For instance, construction sites with empty rooms were difficult to distinguish from each other. Moreover, the movement of humans, who can be temporarily stationary in a room but then suddenly move to a different location, posed a significant challenge for 2D SLAM. In contrast, 3D SLAM could always see the walls and the building at least partly, making it a superior choice for our application.

While practically all SLAM implementations have their own internal representations of the environment, we opted to create and maintain a separate map simultaneously. This was done for a few different reasons. First, the internal environmental representations of SLAM implementations can be difficult to access or utilize and may be in an unusable format or too sparse. Secondly, tight integration between the chosen SLAM implementation and the map used for navigation makes it difficult to switch between different SLAM implementations. Finally, having a separate mapping system allows for deeper customization and integration between navigation methods and the map they rely on.

To address these requirements, we developed a custom probabilistic 3D occupancy grid mapping system, which allowed the robot to build and update a detailed map of the environment. The developed system is similar to OctoMap but utilizes a different chunk-based spatial partitioning method in place of octrees. The 3D occupancy grid was updated with Bayesian updates that consider measurement inaccuracy and existing beliefs of occupancy (prior probabilities). Additionally, the developed mapping system can be automatically initialized with a BIM of the building. This effectively minimizes the need for manual or automatic exploration of the environment that would otherwise be needed to map the environment. The automatic initialization works by converting an IFC file to a point cloud and initializing the occupancy grid based on this information.

Furthermore, we employed deep learning-based 3D semantic segmentation to assign an estimate of the probability that each detected obstacle is static or dynamic. This probability value, ranging from 0 to 1, represents the likelihood of an obstacle being stationary or moving. This deep-learning model was trained using synthetic data, generated using a custom LiDAR simulator implementation within Unreal Engine. By incorporating this semantic information into the occupancy grid, the robot can make more informed path planning and navigation decisions. Additionally, this estimate can be used to effectively remove or segment dynamic and potentially dynamic obstacles from the generated maps.

Figure 5 shows a map of one of the test sites created by the robot. The developed mapping system can create accurate maps of the environment, while the deep learning-based semantic segmentation model can effectively predict which parts of the environment are static. Different colors represent different object classes, with red points representing static obstacles and blue points representing dynamic or potentially dynamic obstacles.

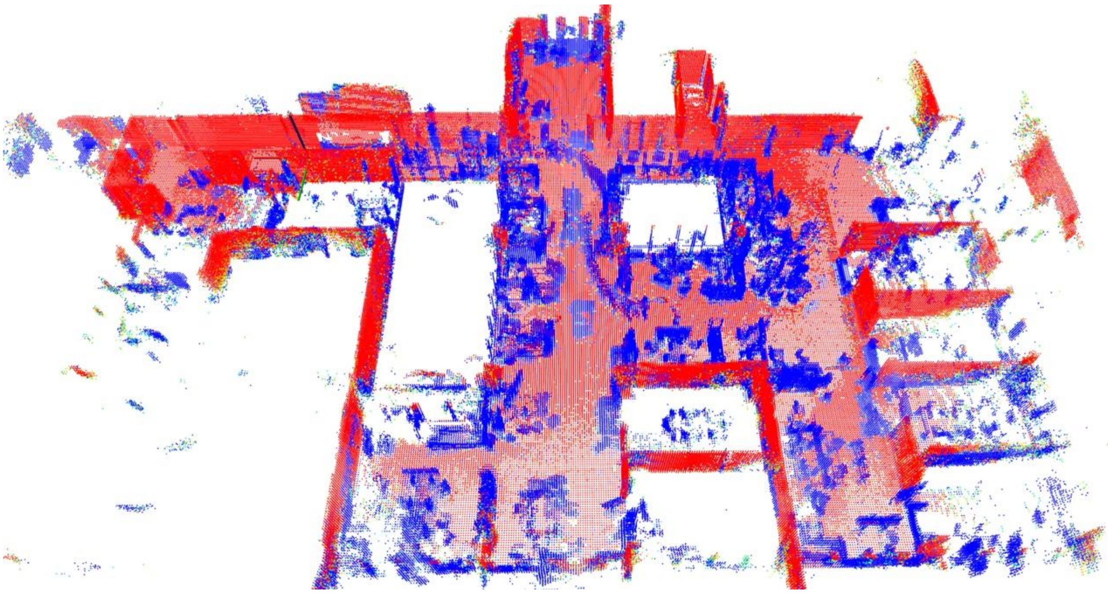


Figure 5: Map of a test site created by Spot. Red points are static obstacles and blue points are dynamic or potentially dynamic points

5.2.2 Planning

To plan paths for the robot to move between data collection locations, we created an A*-based global planner. The path planning algorithm was a modified A* algorithm with a probabilistic cost function. The cost function was designed to minimize the compound probability of encountering obstacles along the planned path. In other words, the path planning algorithm was optimized to ensure that the robot took the least risky path to reach data collection locations. While optimizing for both safety and efficiency is possible, it is a challenging problem. Achieving a good balance between safety and efficiency often requires compromising on one of these factors. However, we note that while we optimized for safety only, the chosen path was regularly also the most efficient.

In scenarios where the path to the goal appeared blocked on the map, the system employed a fallback mechanism to handle dynamic obstacles identified through deep learning-based semantic segmentation. If these obstacles were not directly observed by the robot in real time, they were disregarded in the path-planning process. Since dynamic obstacles can move from their mapped positions, assuming they remain stationary could result in all potential paths being erroneously perceived as blocked. As such, this approach ensured that the robot could find a viable path despite dynamic obstacles.

The continuous path generated by the global planner was simplified to a list of waypoints using a line simplification algorithm (Ramer–Douglas–Peucker algorithm). These waypoints were sent to the Spot robot as movement commands, relying on Spot's internal trajectory following features. The waypoints were sent in order, with a new waypoint being sent as the previous one was reached.

Because developing and testing navigation systems is an iterative process, we opted to utilize an Unreal Engine-based simulator for testing and evaluating the navigation system (Figure 6). The Unreal Engine-based simulator provided a digital environment of the buildings under construction, based on 3D models in IFC format, to simulate the robot's performance in various scenarios. To effectively utilize Unreal Engine in this context, a custom C++ LiDAR simulation component was developed. This component enabled us to generate synthetic point clouds with realistic measurement inaccuracy and noise based on the virtual environment inside Unreal Engine. Further, this enabled the creation of synthetic LiDAR datasets to train semantic segmentation models. By utilizing this simulator, we were able to test and refine the navigation system iteratively, significantly reducing the time requirements and the risks associated with real-world testing.

Figure 6: Unreal Engine-based simulation for testing navigation in the presence of various obstacles

5.2.3 Test Results

The autonomous data collection system was tested in both simulated and real-world environments. Simulation was mostly used to test individual components of the system during development, while real-world tests were used to evaluate the entire solution. The real-world tests featured construction sites at different stages of completion, providing a range of scenarios to evaluate the robot’s capabilities.

In real-world tests, initialization of the system involved creating an initial 3D map based on the BIM model and placing the robot in a pre-defined starting position within the site. The robot's navigation abilities were evaluated through both manual control simulations and fully autonomous navigation tests. These tests assessed the robot’s capability to reach planned destinations, avoid obstacles, and adapt to dynamic changes in the environment, such as moving equipment and personnel.

As the robot traversed the site, it collected data using its onboard sensors. Additionally, the initial 3D map was updated in real time with encountered obstacles. The collected data included not only data for the primary function of autonomous data collection but also additional data that could be used to further refine and optimize the navigation system in an offline environment. The collected data comprised of LiDAR scans, images, and odometry data.

The outcomes of the testing environments were promising. The robot successfully mapped obstacles at construction sites and localized itself within these environments, confirming the system's ability to create and utilize 3D maps initialized with BIM models. Further, the robot demonstrated autonomous navigation capabilities and could reach designated target locations without human intervention. The robot handled unexpected obstacles and adapted its path accordingly. Additionally, the robot automatically collected data that could be used for progress monitoring of construction sites.

6 Ontology extensions

6.1 Requirements for ontologies

The Digital Construction Ontologies (DiCon) (Zheng, Törmä & Seppänen, 2021) provide a basis for representing construction workflow information and integrating heterogenous digital construction data sources. In the context of the ACTOR project, there are new tasks for the ontologies: 1) to provide more detailed representation for specific data sources such as construction images and smart elevator data, and 2) to establish an integrated data lake to connect and serve different automation or AI development tasks. As a higher-level domain ontology suite, the original version of DiCon only provides a generic conceptualization of digital construction workflow. Therefore, DiCon needs to be further expanded to address these two tasks and enable the use cases required in the ACTOR project. In this section, we describe the requirements for ontology extensions.

6.1.1 Requirements of ontology extensions

Need to interpret construction images semantically

Image-based technologies are becoming increasingly prevalent in construction for their cost-effectiveness and ability to capture construction progress. Achieving effective scene understanding requires linking images with background knowledge. Image-based approaches lack interoperability with other construction information systems. Other systems, such as construction schedules, building information models (BIM), sensors, the Internet of Things (IoT), and indoor positioning systems (IPSs), provide contextualized content that complements images. Integrating these diverse sources is essential for a comprehensive digital situation picture that combines the strengths of image-based technologies with rich data from other ICT systems.

Therefore, an ontology should be developed with the following requirements:

1. The ontology should provide a generic and formalized vocabulary to specify the objects and features in a construction image for semantically labeled construction images. This ontology could thus trigger further image interpretation by using feature-based inference with the semantic rules defined based on background knowledge to interpret higher-level image semantics.
2. This ontology should fulfill the missing links between construction entities and image contents, enabling the integration of construction image semantics with other digital information sources.

Need to involve smart elevator data

The vertical transportation data could aid construction management in tracking and predicting personnel and materials movement patterns, optimizing the elevator operation and logistics flow patterns, incorporating robotics onsite movement, and ensuring safety compliance. Integrating smart elevators' data into the digital construction data lake could enhance operational efficiency and logistics planning during construction, providing a seamless interface for monitoring and scheduling elevator use during the construction phase. Thus, smart elevator data should be included in the digital situation picture. This adds the following requirements to ontologies:

1. to precisely and universally provide a formalized vocabulary to describe elevator-related concepts;
2. ensuring semantical linkage between the smart elevator system and other digital construction systems;
3. well-defined relationships between elevator components and their interactions within the construction process. Robust event modeling, incorporating temporal dimensions and sensor integration, becomes essential for capturing dynamic aspects and real-time conditions of the smart elevator system.

Need to provide a structured representation for the construction inference rules

In addition to consolidating information, ontologies can offer structured axioms of domain knowledge with clear definitions and semantics that are compatible with computer processing. These axioms can be converted into a set of straightforward rules that serve as a model for articulating and deducing both explicit and implicit knowledge and information. Consequently, there is a necessity for a platform that encapsulates knowledge of the construction process in a distinctive rule format, thereby simplifying accessibility for intended users. Based on the vision of the ACTOR project, we aim to formulate an ontology that structures the inference rules to support knowledge management and sharing; such ontology should:

1. provide a structured representation to form construction process rules for sharing and reusing;
2. describe the meta information of rules to support the potential user searching for rules based on their demands;
3. provide a computable rule body that can be directly used by the intended users.

6.1.2 Requirements of expanded ontology usage

Need to connect the data with the robotic-based data collection

The integration of robotics into the construction process is not yet widespread. For robots to handle more complex tasks and enhance their utility in the ever-changing conditions of a construction site, they must be capable of adapting to site-specific information and data. A digital situation picture contains a lot of data that could assist robots in navigation and task execution.

Integrating real-time construction information with robotics poses several challenges. Firstly, data from digital systems must be unified into a comprehensive context awareness database, which is necessary for effective robotic navigation. Secondly, developing adaptive context-aware systems requires real-time, accurate sensing and modeling of construction activities. This is complex because construction sites are dynamic. Thirdly, the context-aware system needs automated data processing methods for automatic contextual inferences. Finally, a link must be established between the construction context information and robot programming.

Therefore, in the ACTOR project, we aim to explore how to build an ontology-based semantic digital twin framework that can guide robots in task allocation and navigation, making them responsive to the dynamic nature of construction sites and thereby enhancing their task performance.

Need to act as the data model of construction workflow digital twin

The creation of a digital twin necessitates the harmonization of diverse information technologies. There is a need for well-defined data structures that enhance the compatibility between different systems and contribute to the semantic enrichment of the construction process within a digital twin. These data structures must support various functions, including simulation and analysis. The use of semantic web ontology has been advocated by researchers for developing data models for digital twins (Boje et al., 2020). This method structures domain knowledge and unifies disparate data streams from assorted construction monitoring systems. Ontology, within the sphere of digital twins, presents an innovative method for data management and application.

Most of the existing research on digital twins for construction processes is still in the conceptual phase. Practical implementations that utilize ontology as a semantic framework for digital twins, encompassing data amalgamation, simulation, and interpretation, have yet to be documented. Therefore, this use case will be investigated as part of the ACTOR project.

Need to test its usability with large language model (LLM) advancements

The construction sector is increasingly interested in semantic technologies for improved information management. Ontologies provide an organized way to represent knowledge in the domain, offering a structured model for information that aids in formalizing and combining data from construction workflows. Such a model is beneficial for tasks like data search and logical reasoning. However, crafting SPARQL queries to extract information from data in RDF format can be complex, presenting a hurdle for users without advanced programming skills, thereby hindering the widespread adoption of ontologies in construction.

The current LLMs have capabilities in creating human-like text and formulating SPARQL queries. Therefore, LLMs could be considered to support the retrieval of construction workflow information from the ontology-represented data lake. However, the applicability of LLM needs to be examined using ontologies in the construction field. Such research needs to assess the effectiveness of the capability of LLM to generate the SPARQL queries for construction information retrieval by considering their syntactical accuracy, logical structure, and relevance of the responses they yield.

6.2 DiCon-Semantic Image Interpretation (DiCon-SII)

The DiCon-SII is developed for a dual purpose: (1) to effectively represent and formalize the semantic aspects of construction images, aligning with domain concepts and features, and (2) to establish connections between images and various digital data sources. The DiCon-SII ontology is published in Zheng et al. (2023).

Figure 7 depicts the ontological model of DiCon-SII. All newly introduced concepts and relations within DiCon-SII are identified with the namespace "dicsii:". At the core of the model lies the Image concept, which encompasses two primary aspects. Firstly, it includes the metadata of an image, inheriting metadata-related properties from the DiCon Information Module's InformationContentEntity. For instance, an Image is associated with an ImageFile stored either on devices or in a cloud service. Secondly, an Image is represented with its unique ImageScene, a subclass of the DiCon Context concept. ImageScene describes the detailed contents of the image scene, encompassing objects, their features, and properties. Each ImageScene incorporates a Resource Description Framework (RDF) named graph, providing a comprehensive description of its contents. Regarding interpretation, each image possesses a represented state describable in natural language, inferred from its image semantics and rule-based background knowledge.

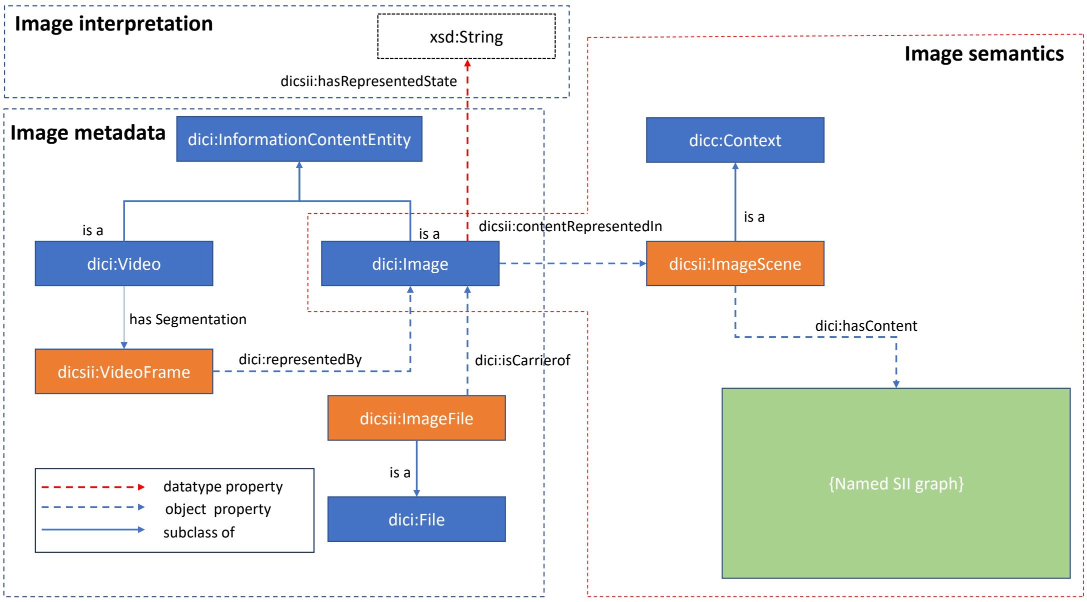


Figure 7. The ontological model of DiCon-SII (derived from Zheng et al., 2023).

Within each named graph, a more intricate ontological structure is employed to encapsulate the essence and attributes of the depicted scene. As illustrated in Figure 8, primary entities such as BuildingObjects, Agents, Equipment, and MaterialBatch are categorized as subclasses of VisibleObject. These entities are key concepts derived from existing literature and are well-documented within the Digital Construction Ontology (DiCon), complete with their inherent interconnections. Consequently, they serve as links to relevant information systems like Building Information Modeling (BIM), Indoor Positioning Systems (IPS), logistics, and Enterprise Resource Planning (ERP) systems. A VisibleObject is characterized by various VisualFeatures, which include Shape, Color, Visibility, and Size, each with specific attributes. These characteristics are also themes of examination in associated studies. Furthermore, there exist VisualPhysicalRelations among different objects, demonstrating their physical interrelations, such as the spatial relationships.

Overall, DiCon-SII was developed to semantically describe formalized and linkable vocabulary for visual construction contents and features. These modeled contents and features enable DiCon-SII to be used to interpret higher-level semantics about the construction image scene and bridge images and other digital systems to help construct an image-involved DTC. More details of this work have been reported in Zheng et al. (2023).

ß

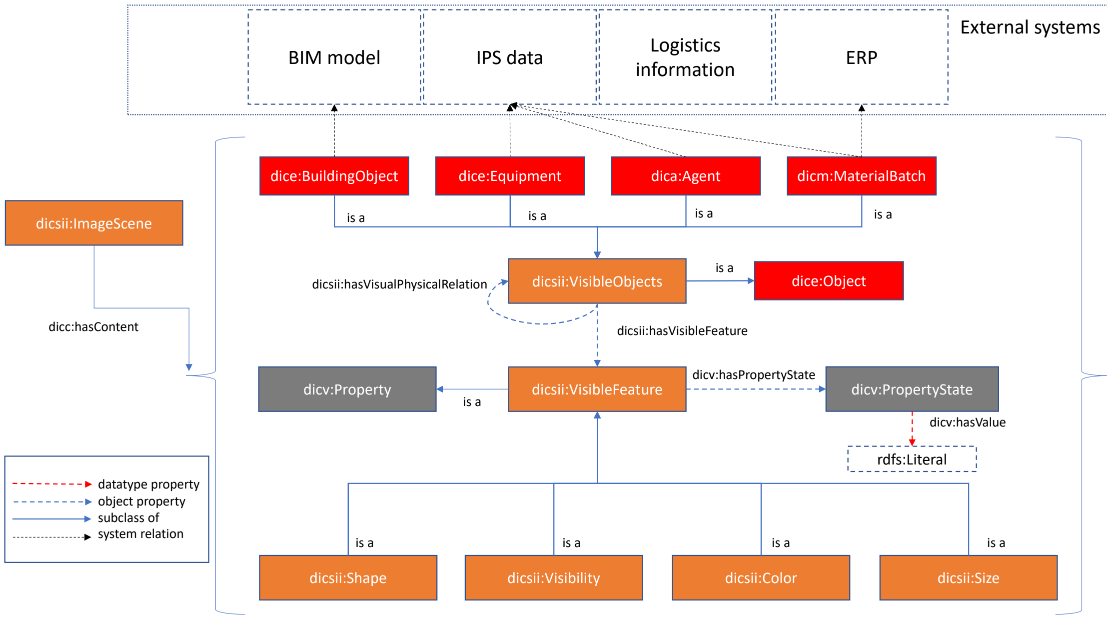


Figure 8. The detailed ontological model for describing construction image contents (derived from Zheng et al., 2023).

6.3 DiCon-Construction Process Library (DiCon-CPL)

The DiCon-CPL aims to model the construction process inference rule that provides a formalized description of the rule and conductible rule body for the sharing and reusing purpose. In the DiCon-CPL, foundational terms for advanced concepts and attributes stem from the DiCon ontology. Additional classes and properties are introduced under the dicr: namespace, complementing the existing structure to articulate rule information via the horizontal segmentation method. The result is reported and published in Zheng et al.(2024). The ontological framework is depicted in Figure 9.

The initial segment focuses on rule meta-information, aiming to formally portray each rule alongside essential background and provenance details. Given the diversity and context-specific nature of rules, the framework opts for a structured narrative of background information over direct classification, thereby aiding users in identifying relevant rules based on meta-information. As illustrated, the model posits that a Rule is crafted by a particular Agent, utilizing a specific DataService for ICT data sourcing, and encompasses Creation Time and Version details. Furthermore, the rules are anchored in source knowledge and discipline, potentially encompassing Constraints to denote precise property values.

The latter segment delves into the rule body, employing SHACL to delineate the inferential rule's structure in an RDF format. SHACL introduces two shape categories: Node Shape and Property Shape. Node shapes impose constraints on subject resources or concept instances of a defined type within the data graph, whereas Property Shapes concentrate on the attributes of classes or their instances. Leveraging the pre-established higher-level Entities and Properties from prior DiCon ontologies, these can be directly targeted as classes or properties within SHACL shapes.

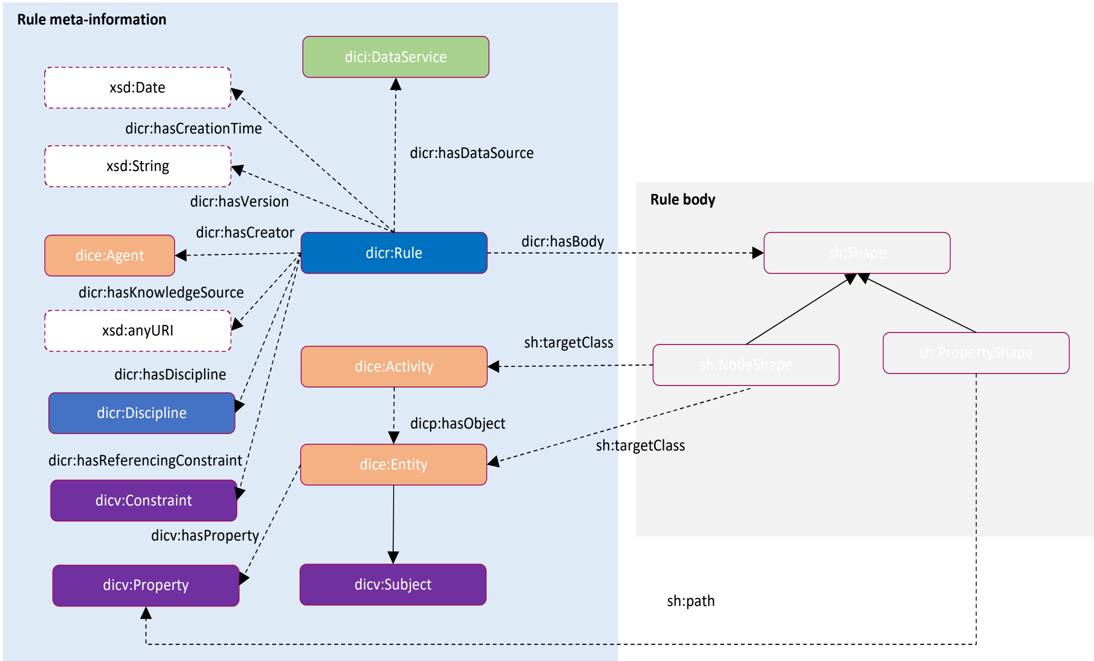


Figure 9. The ontological model of DiCon-CPL (derived from Zheng et al., 2024).

In summary, the proposed DiCon-CPL ontology can provide an explicit description to ensure the rules are shareable. DiCon-CPL also models the meta-information of the rules. In the rule development phase, such meta information would help the developer to formalize the rules to ensure the rule is developed correctly. In the sharing phase, the meta information provides explicit information about the rules the shared users in finding and reusing the rules based on their demands via SPARQL to query the library. Second, building upon the SHACL, the ruling body is conductible via the SHACL inference that can be directly used to infer the process state information, but is also easy to store and share as RDF graphs.

6.4 DiCon-Smart Elevator (DiCon-SE)

DiCon-SE is the extension of the DiCon that involves smart elevator data. DiCon-SE is designed to describe the logistic event data of the elevators. Events include when and where the elevators moved and what kind of material entities or loads they carried/offloaded. This information could support construction operations and logistics management. The development of DiCon-SE used the current version of Kone smart elevator API as a basis and combined it with construction onsite operation and logistics knowledge.

Figure 10 depicts the ontological model of DiCon-SE. The ontological model contains two parts: static metadata and movement tracking data. static metadata describes the metainformation of the elevators and the relations between buildings and sensors. On the other hand, the moving tracking data part centralizes around dice:Movement, which is the class that describes the event of elevator movement. The movement is modelled with different attributes. For example, It is associated with dice:Material by disce:carries indicating material included in movement. Movement also has the temporal data through dice:occupiesTimeInterval to dice:TimeInterval, and spatial data through dice:hasStartFloor and dice:hasEndFloor both connected to dice:Location, as well as load data through dice:hasLoadPercentage to dice:LoadPercentage. Both dice:DirectionState and dice:MovingState are characterized by properties such as dicv:hasValue and dicv:hasTimeofCreation, indicating that each state is defined by a value and a creation timestamp. Overall, the DiCon-SE provides a comprehensive model for tracking the movement of elevators within a building, integrating sensor data to provide a dynamic view of equipment use over time.

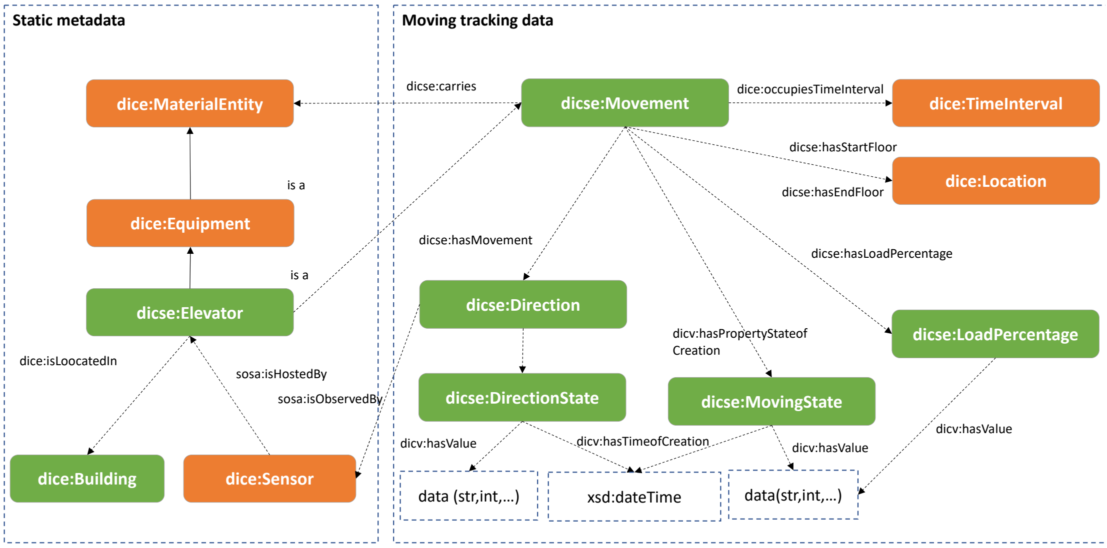


Figure 10. The ontological model of DiCon-SE.

6.5 Ontology implementation

All the mentioned ontology extensions were implemented using OWL encoding to ensure alignment with DiCon and maintain consistency. OWL, a W3C standard language, offers the advantage of expressing richer semantics, representing all concepts as classes, relations as object properties, and attributes as data properties. This OWL-based ontology allows seamless integration with other ontologies in the broader ontological ecosystem. The encoding of all the DiCon extensions in OWL was carried out through Protégé.

7 ACTOR data lake

The DiCon ontologies and their extensions provide a formalized and standardized representation of digital construction data. They can be utilized to establish an integrated data lake. Such a data lake serves as the centered database that enables ACTOR functionalities on the data level. This section describes the ACTOR data lake.

7.1 The architecture of the ACTOR data lake

Figure 11 provides an overview of the ACTOR data lake. This centralized repository allows for the storage and integration of various data streams, including Design information, Schedule, IoT, Logistics, Image-based, and other miscellaneous sources. These diverse data inputs are structured and interpreted within the data lake using the DiCon ontology and its extensions.

The refined semantic data lake can be subsequently utilized by different ACTOR functionalities such as CV processing and simulation. The data lake provides a seamless flow of aggregated data in the data lake to trigger these automated solutions. Additionally, with the semantic rules and the application of LLM tools such as ChatGPT, the semantic data in the data lake can be automatically processed to provide information retrieval to improve the digital situation picture.

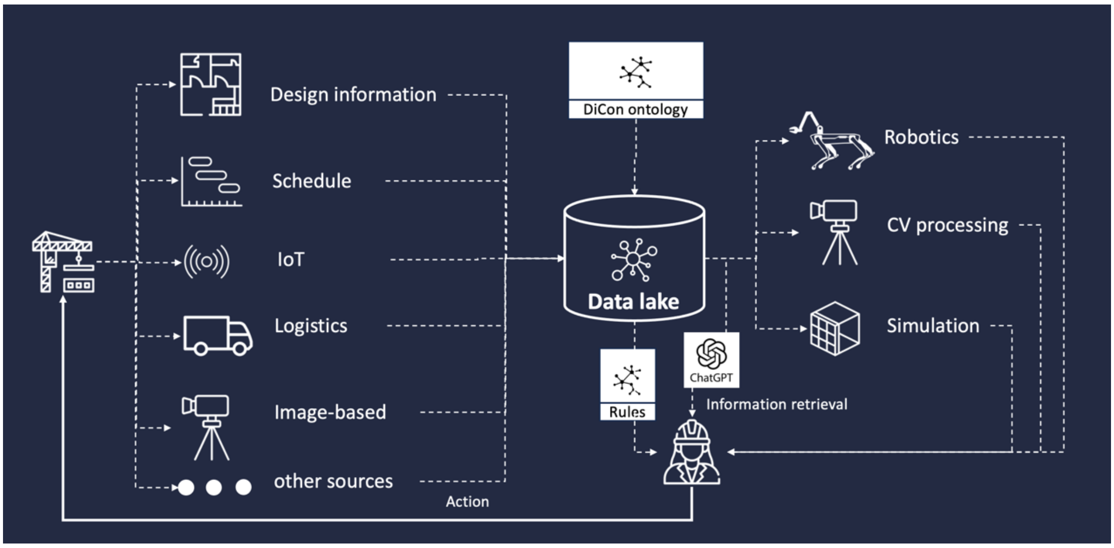


Figure 11. The architecture of ACTOR data lake.

In ACTOR, the software solutions used were heterogeneous and did not directly support the data lake. Data were exported to files or accessed through APIs, depending on source software functionality. The data were then manually edited to conform to DiCon semantic concepts and stored in the data lake. In the longer run, direct connections between source software and the data lake should be implemented. Figure 12 shows an example of how the heterogeneous data from systems are integrated in the data lake in RDF format. In this case, furniture element data was extracted from the Industry Foundation Classes (IFC) file using Python's IfcOpenShell. The subsequent instantiation of BIM data was achieved with RDFlib incorporating IFCOWL and BOT. Concurrently, dynamic IPS data were synthesized in XLSX format and converted into RDF graph. A similar transformation process was applied to the construction schedule and to empirical work rate data. The related instances in these graphs were aligned based on their interrelations. For instance, using the location to map locations in the different data sets.

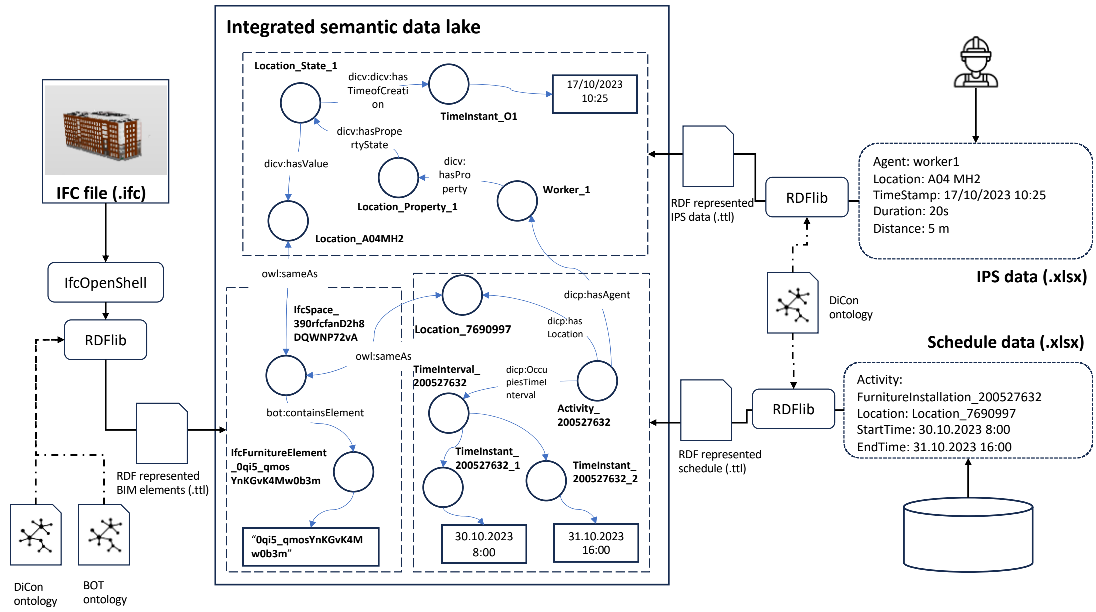


Figure 12. Example of integrating the different data into data lake.

7.2 Applications of ACTOR data lake

7.2.1 Semantic Digital Twin for adaptive robotic construction site navigation

The data lake could aid in the robot's automated inspection planning. Data lake can enable autonomous navigation by identifying locations and tasks that should be inspected. We developed a framework called Digital Construction Robotic Context Awareness (DiCon-RCA). DiCon-RCA system's structure is depicted in Figure 13. This framework harnesses dynamic data from the physical world, capturing the indoor construction process through various digital systems such as Project Status Information (PSI) and feeding it to the Digital Twin (DT) in real time. Concurrently, it gathers static Project Intent Information (PII) from the construction project as additional input DT (Sacks et al., 2020). These two types of data are stored in the ACTOR data lake and form the semantic DT database visualized as an RDF graph.

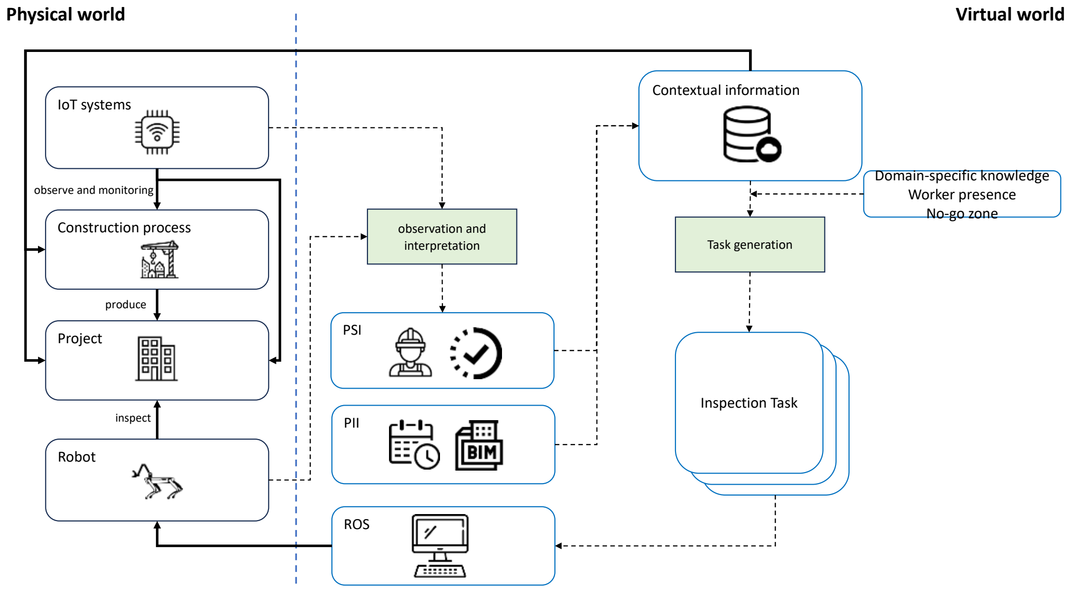


Figure 13. DiCon-RCA framework.

As shown in Figure 14, accessing the DT database realizes context awareness of the indoor construction process through a series of rule-based inference mechanisms. These mechanisms consist of semantic rules designed to deduce the context as an input for the navigation system. To ensure the DT data model remains an accurate digital reflection of the physical space, the robot updates the database with newly collected information and data.

The task generation and location identification process begins by using SPARQL queries against the semantic contextual database to identify ongoing tasks during the field test's specific timeframe. For instance, the current task might be furniture installation based on the construction schedule. Consequently, the inspection location range is set to monitor the progress of the furniture installation task on the specified floor. Using this data, the next step involves pinpointing precise locations and specific furniture items being worked on where worker presence is observed. Based on Zhao et al. (2019), the detection of worker presence serves as an indicator of actual value-adding activities. Worker presence is used to verify the locations of actual construction work. This identification is facilitated by retrieving the corresponding GUID of the location from the BIM model, with the relevant SPARQL queries. When planning the navigation route, the robot can include the resulting locations as data collection destinations.

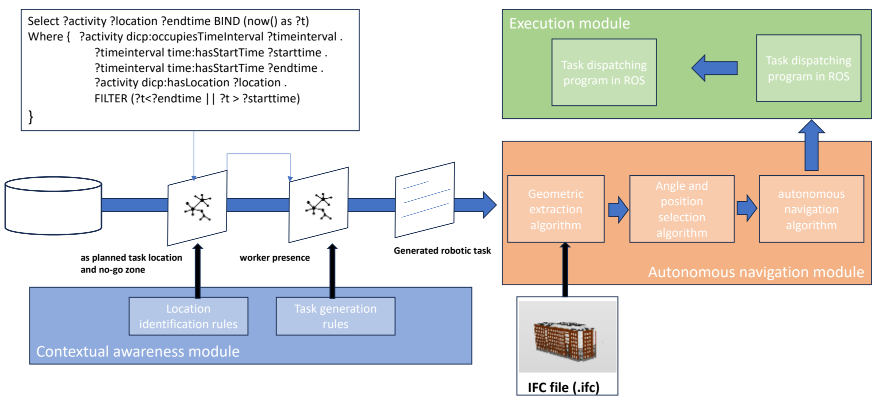


Figure 14. Workflow of DiCon-RCA.

7.2.2 Semantic Digital Twin for simulation

Several scholars suggested the use of semantic web ontology to develop data models for Digital Twins (DT). Most existing studies on DT in construction are still conceptual. Therefore, we developed an ontology-based method to create a semantic DT. Our approach integrates ideas from Digital Twin Construction (DTC) (Sacks et al., 2020) and Construction Digital Twin (CDT) (Boje et al., 2020), culminating in a defined framework known as DiCon-DT. The DiCon ontology enables DiCon-DT to collate and represent various data of indoor construction processes effectively. DiCon-DT features a projection system to simulate indoor construction scenarios. DiCon-DT was applied and tested in a real-world case study focusing on the furniture installation process in an actual project. The assimilated data feeds into an Agent-Based Model (ABM), replicating worker movements based on worker presence (Zhao et al., 2019) and incorporating empirical data on worker behavior to estimate work progress. Additionally, DiCon-DT semantically interprets simulation results through rule-based inference, enabling automated simulation analysis.

Figure 15 depicts the structure of the DiCon-DT system. The physical world supplies the DT with real-time, dynamic data from various digital systems concerning the indoor construction process. Additionally, static information from the construction project is gathered and fed into the DiCon-DT. Beyond the physical realm, the DiCon-DT encompasses three additional layers: the semantic DT layer, the situational awareness layer, and application layers. DiCon considers the integration of different digital data sources of each flow of the construction process. The integrated data is called the semantic DT database and is represented as an RDF graph.

The situational awareness layer plays a pivotal role in DiCon-DT that functions as an integration point for simulation and interpretation, enabling them to access the integrated DT database. The raw data and simulation results are serialized into the RDF DiCon. Semantic rule-based inference can then be used to achieve an automated interpretation of indoor construction process situation. Rule-based inference is an important application of the semantic web to allow the computer to automatically reason based on the data and generate new information and knowledge. To keep the DT data model always a digital mirror of the physical space, the information in the DT data model is updated according to the update rules. Experiments related to agent-based simulation using the integrated DT database are described in more detail in section 9.3.

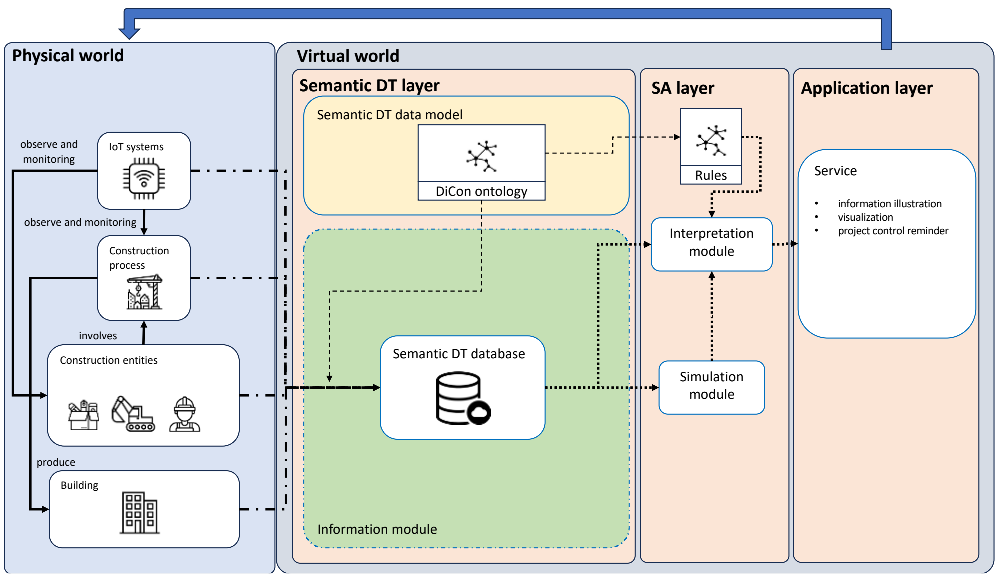


Figure 15. The DiCon-DT framework.

8 Improving digital situation picture with computer vision

Automatic data collection with robots collects mainly visual data of objects or locations of interest. The visual data can be stored and linked to other data using DiCon-SII. However, to be useful for digital situation pictures, in addition to storing the image, it should be analyzed and used to update other data streams. Construction progress and identifying materials are particularly relevant for digital situation pictures. In ACTOR, we focused on drywall progress and material detection.

8.1 Detecting drywall progress

The general method of construction progress monitoring consists of manual observations and reporting the data in weekly contractor meetings. The manual observations are subjective, as the reported data could be incomplete and not clearly defined (Zhao et al., 2021). Automatic construction progress monitoring is desired to solve the problems with manual data collection. The construction scene is dynamic, immersive, and crowded, which renders the interpretation of the visual data collected complex (Fathi et al., 2015). The geometries and materials change frequently during construction (Paneru & Jeelani, 2021). Also, there are variations in a scene, like light variations, the presence of dust particles & reflective objects, which may affect the accuracy of the data collected on-site and consequently the interpretation (Mirzaei et al., 2022).

For construction progress monitoring, attempts have been made to automate the progress assessment based on work package assessment & CV, by one-to-one comparison between as-built and as-designed data (Yang et al., 2015, Roh et al., 2011, Hamledari et al., 2017). In most works, the use of CV is limited to object classification, and comparison between objects is done visually in as-built and as-designed data. (Masood et al., 2020). The approach of comparing as-built to as-designed with occupancy-based methods does not address instances where a single element of construction goes through different stages, such as drywall.

The initial step in automatic progress monitoring is the detection of objects on the construction sites. The detection of objects can be achieved by deriving features by using traditional image processing, but that is application-specific. For a more holistic approach, a generalizable solution is more desirable, which can be achieved through deep-learning approaches. For deep-learning based solutions, there is a requirement of algorithms to be trained on construction site annotated images.

Our aim was to have a holistic approach that can detect all the as-built stages of drywall to improve the digital situation picture on construction sites. Therefore, we chose the stages of drywall so that all the prominent steps are covered as follows: Stage 0: No drywall; Stage 1: Installation of studs; Stage 2: Gypsum Paneling; Stage 3: Electrical and plumbing works; Stage 4: Insulation work; Stage 5: Wall-closed/only paneled side visible without studs; Stage 6: Plastering of the wall; Stage 7: Painting of the wall. We formulated our problem as a computer vision task, specifically object detection, which amounts to the detection and location of objects of interest in an image. We used a deep learning approach in combination with transfer learning to train a model for object detection (Pan & Yang 2010). Transfer learning methods allow leveraging already obtained knowledge for solving different but related tasks.

The data was initially collected as 360-degree videos on two construction sites. Drywall snapshots were taken from these videos to generate images for each stage of the drywall installation process. Generally, it was decided that about 150-200 images of each stage would be enough to change the parameters of the pre-trained neural network. The images were annotated in a software called Image Annotation Lab, which was selected after a comparison of several other systems (VGG annotator, MakeSense, RoboFlow, and Supervisely) based on parameters such as easy GUI, price, and output formats provided.

The labelled images were then fed to YOLOv7 and v8 deep learning models, and transfer learning was done. The models yielded good results on images that were not in the training set, and all the stages of drywall were successfully detected. Some of the results are shown in Figures 16-18 below.

Figure 16: Detected Studs and Paneling stages of drywall.

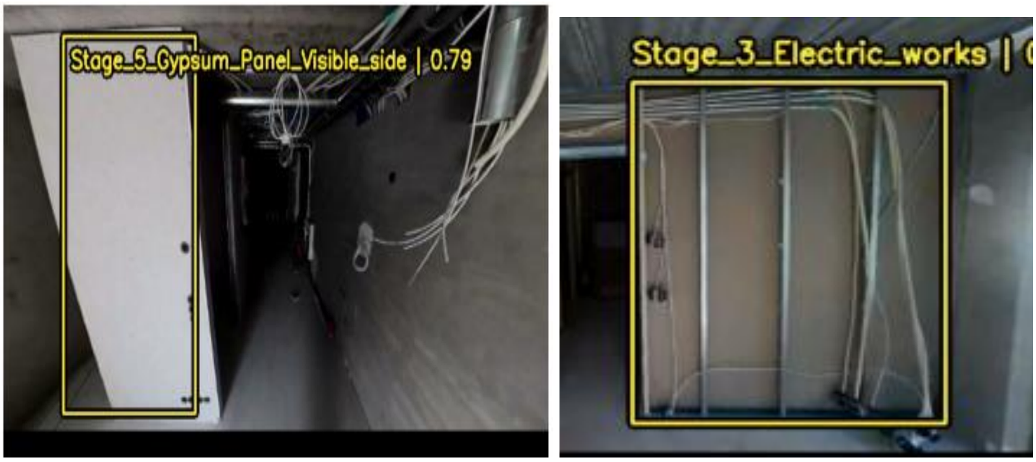


Figure 17: Detection of visible side of gypsum board and electrical work stage

Figure 18: Detection of Plastering and Painting & Finishing

8.2 Detecting drywall materials

Materials are one of the important flows that are needed for construction. In drywall case, materials include metal studs and gypsum boards. We used computer vision to identify the various material required. For this purpose, we used screenshots from 360-degree videos and used them for semantic segmentation. “Segment anything”- model from Meta was used to segment the images and identification of material was done manually. In Figure 19 one result of the segmentation is shown.

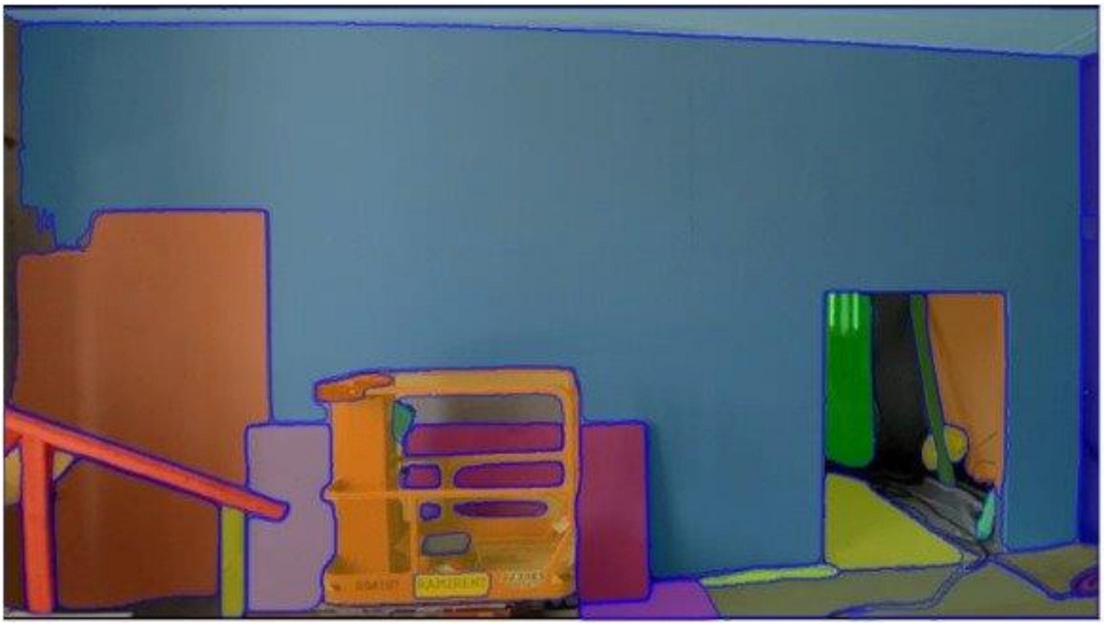


Figure 19: Semantic segmentation of drywall and other equipment and materials near dry wall

Based on data collection activities in two projects, there was not enough training material to reliably detect drywall materials. There are many walls in a construction project continuously progressing from one state to the next, so there is enough data for progress detection. There were much fewer material storage areas in the projects we examined. Therefore, transfer learning could not be performed on a very limited dataset. Yet, the observations in these projects helped us plan an approach towards material tracking in future projects. Material map of the construction site needs to be created to make it convenient for the workers to locate and use the right material for a task.

9 Agent-based simulation models as foundations of digital twins

The working assumption of the project was that construction schedules are not accurate enough to serve as digital models of the construction process. It was envisioned that agent-based simulation models could be the basis for digital twins, which could enable continuous optimization on the worker level so that instructions could be delivered visually using digital visual management solutions. In the project, we developed a proof-of-concept simulation model. Unfortunately, it was impossible to properly validate the digital twin because computer vision algorithms were developed to detect drywall progress, but drywall workers did not want to participate in the study. Therefore, the concepts were validated using a different trade (furniture installations) without all the data streams available. This chapter describes the simulation model and tentative tests of the digital twin.

9.1 Agent-based simulation model of drywall installation

This section provides an overview of the fundamental elements of ABM for drywall installation process developed in the project. The elements of the model were selected based on previous research on worker productivity. The goal of the model is to model worker behavior accurately enough that it can be used to propose the next task of the worker. Additionally, it can be used to evaluate different scenarios of site layout.

We have utilized IfcOpenShell to extract BIM data, which we then incorporated into our agent-based model grid. This integration allows for precise visualization and simulation of the building's spatial layout and operational zones, such as apartments and takt areas. The detailed grid system, with each cell measuring 0.5m x 0.5m, enables high-resolution tracking of worker positions and activities in real time. By mapping out various structural elements like corridors, rooms, walls, doors, and staircases, the grid system accurately depicts the workers' locations and interactions with the building environment.

This approach enhances the understanding of movement patterns and traffic flow, thereby improving the overall efficiency of the simulation. The model represents the construction site and agents at each time step. The model advances in steps. The state of each agent is evaluated in each time step. We defined multiple types of agents in the model to represent the various entities involved in construction and furniture assembly: workers, zones, and furniture BIM elements. Each agent has attributes and specific behaviors. Figure 1 shows a simplified illustration of agent attributes, a sample of their states, and system KPIs.

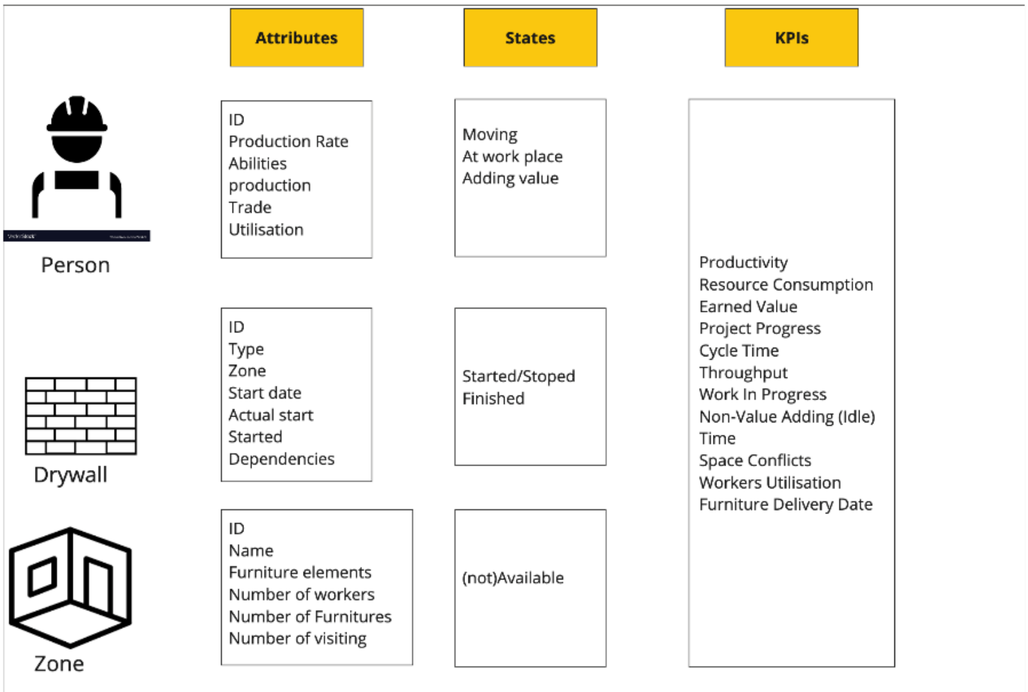


Figure 10. Main agent's parameters, states and KPIs used in the simulation

The worker agent is a construction worker that belongs to a trade and can perform specific tasks with specific production rate. The main worker attributes are described in Table .

Table 2 Attributes of Worker Agent

Attribute	Description
ID	A unique identifier assigned to each worker agent.
Current Position	X and Y coordinates of worker
Production Rate	Represents the worker's productivity in completing tasks.
Abilities	Captures the tasks the worker agent can perform
Trade	Refers to the trade of the worker agent belong to.
Utilization	Percentage of working time worker was in working zone
Worker Times	Arrival, leaving, and break times determined by triangular distribution function (TDF)
Probability density function (PDF) for durations in (non)working areas	Probability density functions are unique for each task type, that best fits the observed real-time data from IPS. They are used to determine frequencies and durations spent in working and non-working areas for each task type.

Figure (21) illustrates a flowchart depicting a worker's actions at a particular site. The process of worker arrival comprises multiple stages. Initially, it commences by verifying whether the worker has reached the site during the specified time step. If the worker has arrived, the flow proceeds; if not, it returns to the beginning, indicating the worker's absence. Upon reaching the site, the worker establishes their intention, indicating their intention or task at the site. This intention could pertain to work, work-related support activities, or any other specific purpose. Prior to gathering actual data from the site, intentions are formulated based on prior studies (e.g., (Görsch et al., 2022)). Subsequently, once indoor positioning data is accessible, the probability density function is computed using the data.

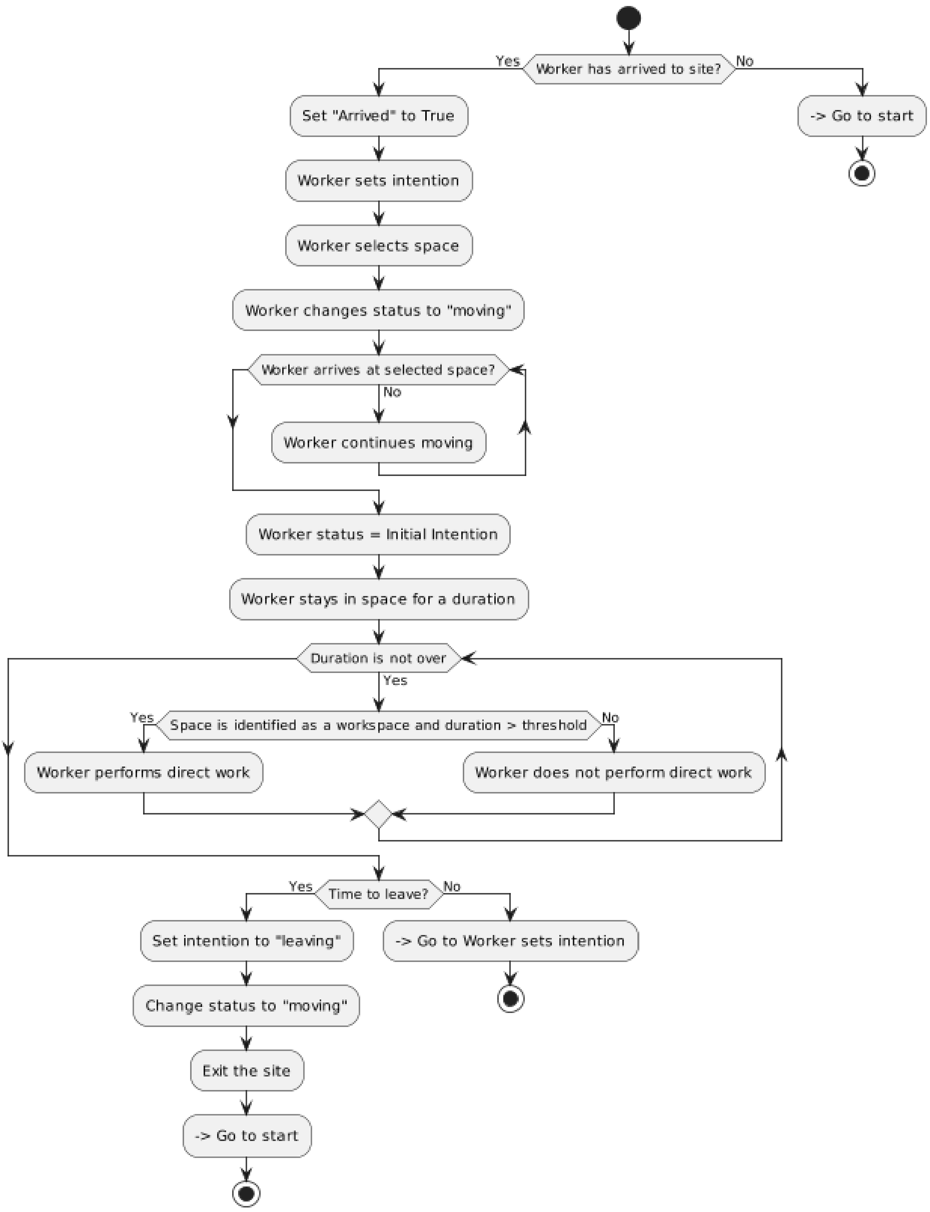


Figure 21 Worker process chart

The worker chooses a particular location within the site based on various factors, including the sequence of the initial plan and the worker's current task. Subsequently, upon selecting the space, the worker changes their status to "moving," signifying their movement towards the designated area. The worker aims to avoid obstacles and identify the most efficient route by utilizing a pathfinding function (A* as explained in the previous section).

Upon reaching the designated space, the worker adjusts their status to align with their original intention and executes the specific task associated with that location. It is assumed that direct work is carried out within the allocated space. Tasks related to other areas may include gathering materials, seeking clarifications, obtaining specific tools or equipment, or taking breaks in designated rest zones for lunch or coffee. Workers typically arrive at the site at the beginning of their shift and depart at its conclusion, navigating from their current position to the site's entrance or exit. The probability of visiting each area and the duration of stay are initially established through prior research and subsequently adjusted based on empirical data regarding movement frequency, duration, and presence within the indoor positioning system. As the various locations shift within the dynamic site layout, workers adjust their movements accordingly (e.g., proceeding to the material storage area irrespective of its current position).

Effective congestion management is a critical aspect of this system and is classified into two primary categories. The first type pertains to instances where non-assigned workers visit a specific zone, resulting in unnecessary congestion. The second type occurs when the number of workers assigned to a task within a zone surpasses the optimal capacity for that activity. Furthermore, certain areas limit the maximum number of workers allowed to prevent overcrowding.

The system determines the end of a worker's presence in a specific area. When a task is completed or the selected time elapses, the worker updates their status to "leaving" and changes back to "moving." They then utilize the pathfinding feature to progress toward the subsequent destination, thereby beginning a new operational cycle.

BIM Elements, which symbolize distinct work elements, are characterized by attributes such as positions, zones, tasks, workload per task, finish-to-start dependencies, statuses, start and end dates, actual start and end dates, and progress rates. These entities engage with other components in the simulation, including workers and other BIM elements. Zones, defined sections within the construction site, possess characteristics like locations, BIM elements contained within, and workers present within. Zones interact with other agents, contributing to the dynamic essence of the simulation. Walls, as physical obstructions on the premises, serve as barriers that influence workers' movement. Workers navigate through unobstructed pathways, avoiding walls and other hindrances as they transition between areas. This level of modeling detail is essential to ensure the simulation outcomes are relevant to the workers.

In addition to attributes, agents exhibit behaviors and interact with one another. For instance, in the context of drywall installations, Figure 22 illustrates various interactions within the model. These interactions are broadly classified into three primary categories: construction worker-drywall element interaction, construction worker-zone interaction, and drywall element-zone interaction.

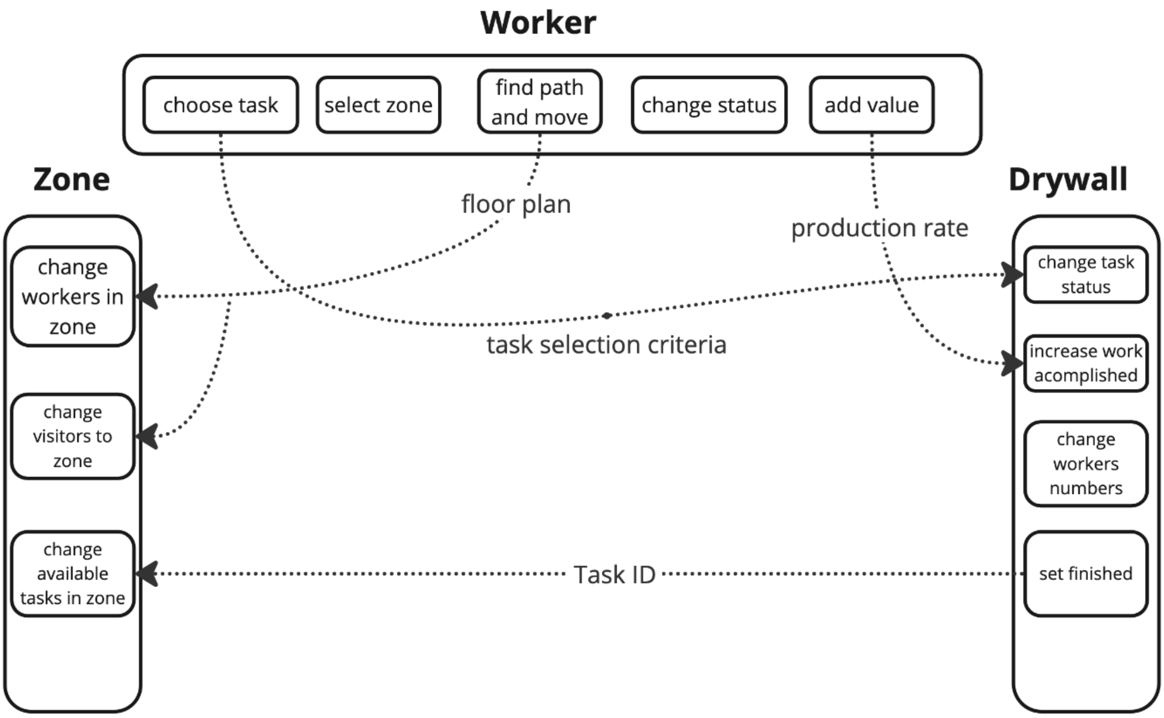


Figure 22 Example of agents behaviors and interactions.

The interaction between construction workers and drywall elements entails the direct involvement of construction workers with drywall elements. Various factors influence these interactions, such as the workers' skills, production rate, workload, and specific tasks. Construction workers carry out their tasks on drywall elements.

Interaction between construction workers and zones involves the movement of construction workers across different zones within the construction site. They transition between working and non-working zones to access resources, acquire information, complete production tasks, and take breaks. The presence, proximity, and suitability of zones significantly affect the movement and resource utilization of construction workers. Well-structured and designated uncongested zones facilitate a seamless workflow and boost worker efficiency.

The interaction between drywall elements and zones encompasses assigning and positioning drywall elements within specific zones based on their characteristics and needs. Zones within the construction site offer the physical space for installing and organizing these elements. For instance, specific zones may be designated for storing drywall materials and tools. Zones can be aligned with the site's location-based construction schedule.

9.2 Empirical tests

During the data collection process, the project was in the midst of the interior construction phase, focusing on the execution of furniture installation tasks. The drywall workers refused to consent to location tracking due to the shift from drywall to furniture installations. Consequently, integrating other system components, such as computer vision for progress evaluation and visual instructions for workers, was not feasible during the testing phase. Below, we present an overview of the available data sources and field-level data acquisition:

1. IPS data: Within the construction site, the principal contractor implemented the Noccela system (Noccela Oy, n.d.), an Ultra-Wide-Band (UWB) tracking system, to monitor indoor installation workers (e.g., furniture assemblers, laminate installers, and haulers) with their explicit consent. This indoor positioning data captures the movements of workers throughout the site. The IPS data, formatted in Excel (xlsx), was furnished by the contractor, as direct access to the IPS Noccela database was not granted. Prior to transmission to researchers, the contractor processed the data, eliminating irrelevant information. In the context of a proof-of-concept case study aimed at system evaluation, a total of 7192 indoor position records were collected, each representing a construction worker's presence at a specific location along with their corresponding start time and duration. An illustration of the IPS data is presented in Table 3.

Table 3 Example of IPS data from Noccela system in Excel format.

Agent	Location	Start time	Duration
KALUSTE 1	A06	2023-10-20 07.10	0h 0 min 2 s
KALUSTE 2	A10	2023-10-20 06.05	0h 0 min 9 s
KALUSTE 3	A11	2023-10-20 07.11	0h 0 min 49 s
KALUSTE 1	A12	2023-10-20 06.50	0h 0 min 22 s

2. The architectural Building Information Model (BIM) was acquired from the contractor in the Industrial Foundation Class (IFC) format (.ifc). This particular model corresponds to a Level of Development (LOD) 300, encompassing a comprehensive detailing of all interior construction elements.

3. The Finnish productivity database Ratu (Card 0421) (Rakennustieto, 2023) was utilized in the empirical analysis to establish the baseline work rate of laborers. The recommended work rate outlined in the manual was extracted and transcribed into an Excel file (xlsx), as delineated in Table 4.

Table 4 Empirical work rate Ratu Card 0421

Furniture type	Working part	Work rate (h/man*piece)
Table cabinet	table cabinet compilation and installation	0.20 + 0.30 = 0.5
Wall cabinets	wall cabinet compilation and installation	0.20 + 0.5 = 0.7
Closet	closet pantry cabinet compilation and installation	0.25 + 0.3 = 0.55
kitchen equipment	Including sink, other equipment, and legs.	0.5 + 0.35 + 0.28 + 0.1 = 1.23
kitchen furniture ensemble	2.5 to more than 6.5 m2	Between 10 and 20
other appliance installation	Racks, cloth room racks, cloth closet	0.47 + 0.20 + 0.50 = 1.17

4. Construction Schedule: The construction schedule offers comprehensive details regarding the construction procedure, encompassing task names and IDs, task types, task locations, anticipated start and end times, and task locations. The weekly schedule was obtained from the principal contractor through Microsoft Excel.

9.3 DT and Simulation implementation

The simulation model was executed utilizing the open-source Python framework provided by MESA (n.d.). MESA's structure emphasizes modularity, offering elements for modeling, analysis, and visualization. This adaptability enables seamless integration and expansion of elements to formulate diverse model types. Figure 23 delineates the workflow and constituents of the simulation execution.

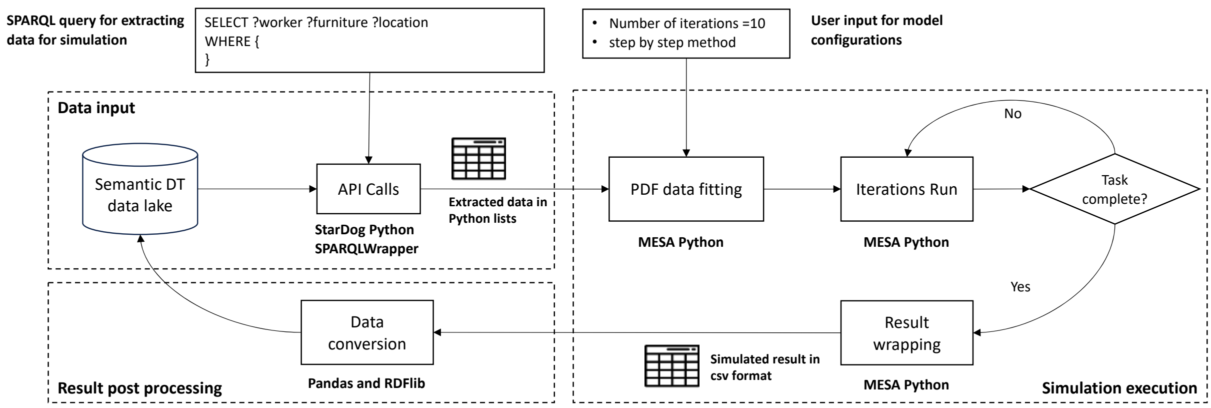
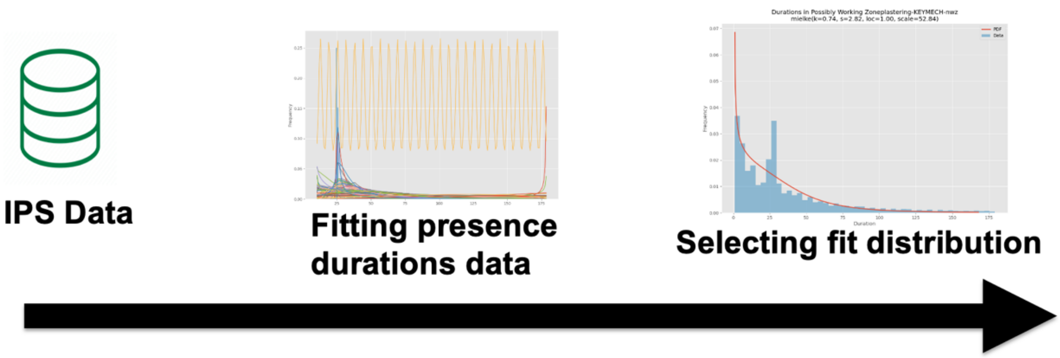


Figure 23 The workflow of the simulation implementation

The initial phase involves data input, where semantic information is extracted from the data lake to initiate the simulation model. This step utilized the StarDog Python API for executing SPARQL queries to fetch data in Python list structure aligned with the simulation model attributes. These attributes consist of task details such as location, worker count, start and end times sourced from the as-planned graph; furniture specifications and quantities from the as-design graph; work rate data from the default graph; and IPS data reflecting worker presence on the 4th floor from the monitoring graph. Subsequently, the initialization algorithm decodes the received data to populate the model with agent instances and configure parameters for each simulation iteration.

Examining IPS data pertaining to workers enables the determination of probabilities associated with their movement between working and non-working zones by counting occurrences within a specified timeframe. Additionally, this data aids in approximating the presence duration within each zone by aligning the IPS data with 104 established Probability Density Functions (PDFs) and selecting the most suitable PDF to determine presence durations within zones, exemplified in Figure 10. The Alpha PDF is described by three vital parameters: shape (a), location (loc), and scale. The shape parameter (a) defines the overall distribution shape, where a value of 0.00 signifies a specific form within the alpha distribution family. The location parameter (loc) shifts the distribution on the x-axis, with a value of 150.52 indicating its center. The scale parameter influences the distribution spread, with a scale value of 1152.452 dictating the data point dispersion level. These parameters collectively (a = 0.00, loc = 150.52, scale = 1152.452) establish the characteristics of the alpha PDF, outlining its shape, position, and scale concerning the dataset under scrutiny.



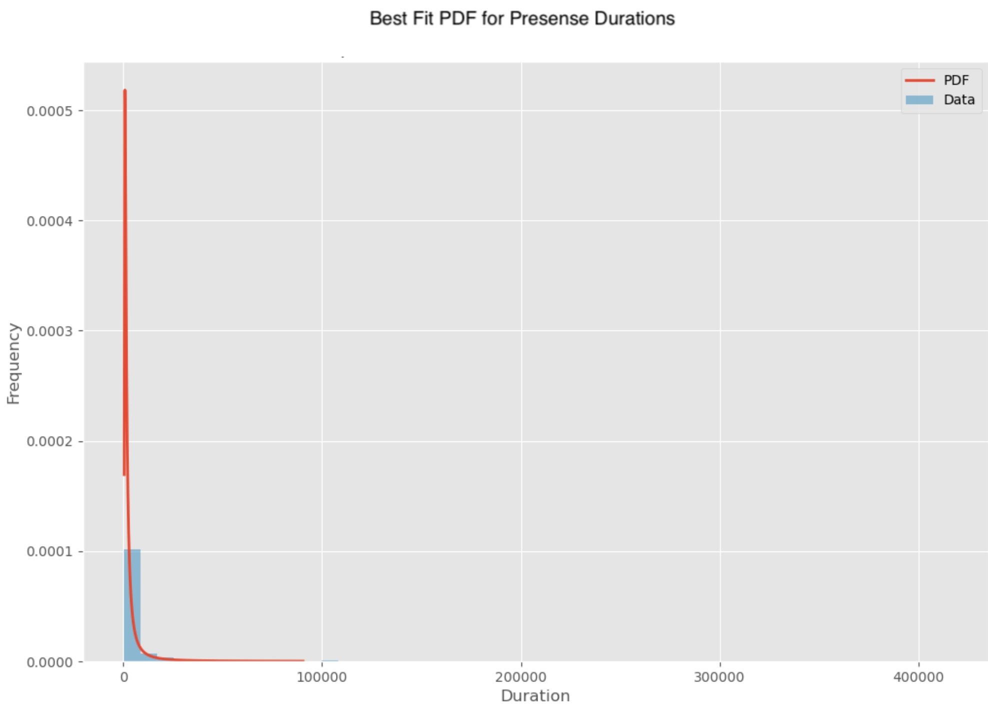


Figure 24. *Best-fit PDF distribution based on workers IPS data*

By integrating the work rate and quantity data with the most suitable Probability Density Function (PDF), we made an estimation of the task progress on the 5th floor, mirroring the observed results on the 4th floor. The estimation process followed the methodology of worker's presence as outlined by Zhao et al. (2019). Ten simulations were executed, generating outcomes encompassing workers' movement characteristics, status updates of various zones and furniture components at diverse time intervals within the simulation. Additionally, it provided an estimation of the time needed to complete tasks in each apartment and floor in a tabular layout. A selection of outcomes is visually represented within the box plot in Figure 25. Subsequently, the simulation results were transformed into a Resource Description Framework (RDF) graph linked to the ontology network utilizing the RDFlib and Pandas Python libraries. This RDF graph, which represents the simulated scenario, was transmitted back to the data repository for in-depth analysis, enabling a comparison with the scheduled data to facilitate progress assessment.

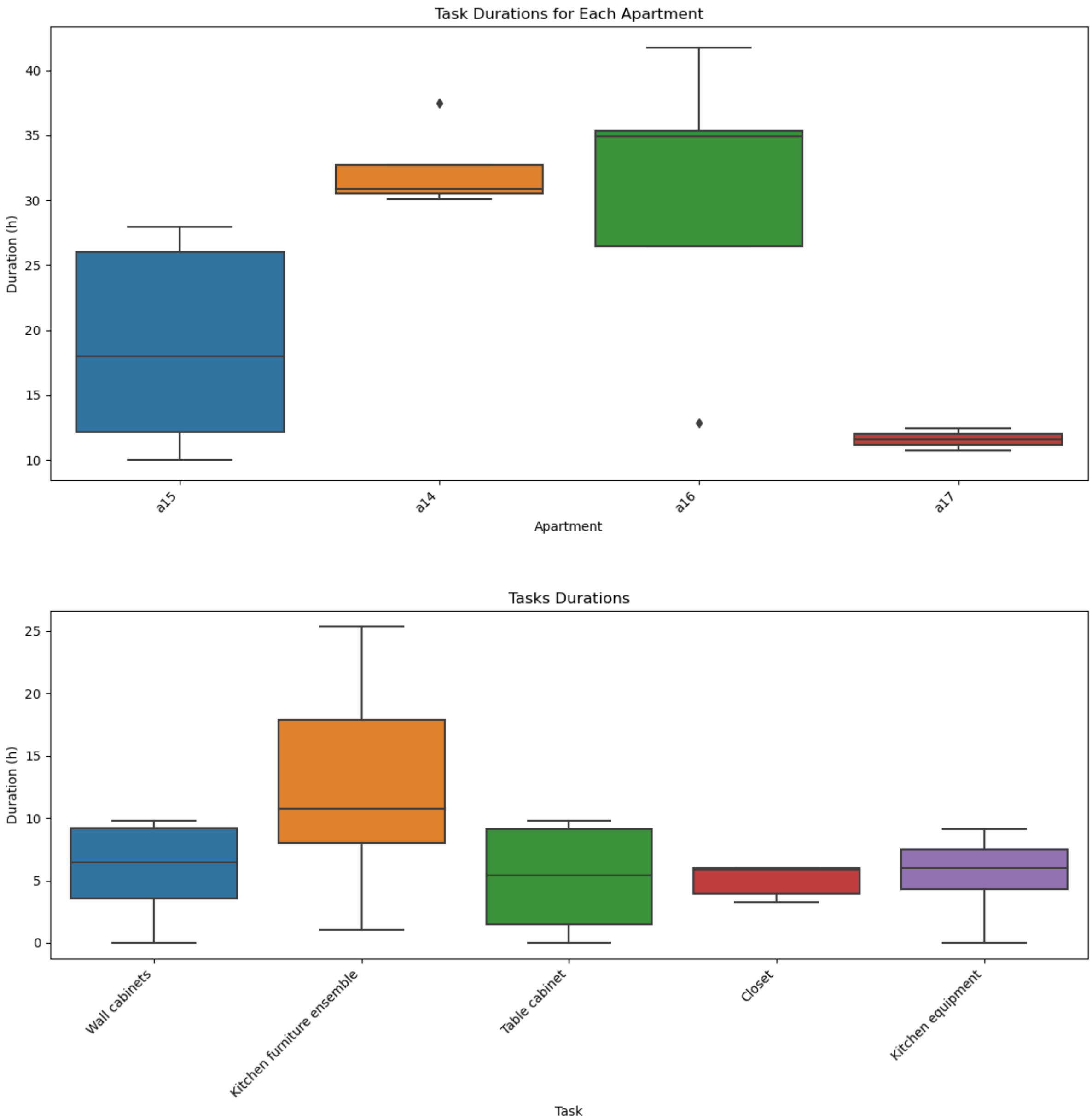


Figure 25. The Durations of Furniture Installation in each apartment (top) and for each task (bottom)

The simulation model was able to predict the durations of upcoming tasks. It could be used to evaluate several interventions, such as increasing the number of agents, moving the locations of non-work areas (e.g., by moving storage areas closer), or taking managerial actions to decrease the amount of movement. However, because the task where IPS data was available (furniture installation) was different from most of the data in the data lake (drywall installation), full validation of the simulation model had to be postponed to future research where the intention is to get a fully aligned data set.

9.4 Modeling worker decision making in DT

Workers make decisions differently. Some use the official schedules while others make their own plans (Görsch et al., 2023). Because each worker has their own preferences, it is impossible to develop a generic utility function that would apply to all workers of a trade with the same weights. However, it may be possible to come up with ways to arrive at a decision, which will be weighed based on worker preferences. In this project, we attempted to identify which factors impact worker decisions. Ultimately, the digital twin system could provide recommendations tailored to individual's preferences. Simulation could use the same preferences (if known or if learned) and otherwise be based on observed behavior in the past. Based on workshops with consortium members, we came up with the following factors:

- Size of work – preference to do small or big tasks
- Location of work – with respect to material storage and the current work location
- Priority – critical for other trades or opens up more work for the worker
- Congestion in work location – critical for labor productivity
- Time needed to make space ready for work

All these factors can be sensed using different data streams. The size of the work is obviously known based on the design. Work location requires knowledge of material storage (from robotic data collection and computer vision) and current location (From indoor positioning). Priority is defined based on the schedule and dependency relationships of elements. Congestion can be sensed using IPS. The time needed to make space ready for work could be sensed by implementing a computer vision model trained to detect the readiness of workspace. Thus, the digital twin system envisioned in this project could enable ensework status and propose the next task based on worker preferences. Potentially, multiple options could be presented in visual management by showing the optimal task based on priority and the optimal task based on personal preferences. A conference paper published based on this work elaborates this concept further (Abou-Ibrahim et al. 2024).

Visual management presented in the next chapter depends on optimizing the next activity for the construction worker. The work should be doable and "logical" for the worker. The concepts were simulated (Abou-Ibrahim et al. 2024), but we could not complete a practical test during the project time frame because drywall installers opted out of the research, and other work packages used drywall as the use case.

10 Digital Visual Management assisting Construction Actors Decision-Making

The final technical work package of ACTOR was related to digital visual management on the worker level. The instructions optimized based on a digital situation picture should be shown to workers at their work site. The work package included identifying current challenges of VM in construction, interviewing workers to identify their challenges, refining requirements for digital visual management on worker level and developing and testing to digital visual management solutions.

10.1 Current Understanding and Challenges of VM in Construction

The management and research perspective on (digital) visual management ((D)VM) in construction reveal strategies at an early stage, driven by individual initiatives rather than company-wide approaches. Despite extensive research on digitalization and VM, workers' viewpoints are often overlooked, leading to a focus on managerial decisions (Reinbold et al., 2022). Deficiencies in visual information dissemination to site personnel pose challenges, exacerbated by resource constraints hindering timely process information for workers' awareness (Tezel and Aziz, 2017). Current VM device placement in site offices, not near job sites, limits optimal use (Tezel et al., 2011; Reinbold et al., 2022). The complex relationship between VM and ICT is inadequately understood, hindering communication facilitation (Pedo et al., 2022; Reinbold et al., 2019). Limited perception of VM as a production control tool hampers broader adoption (Tezel et al., 2010). This research emphasizes the need for a comprehensive approach, integrating both managerial and worker perspectives for effective VM implementation.

From the workers' perspective, challenges in VM emerge as they grapple with outdated information, affecting situational awareness. Analogue VM devices catering to site management contribute to information silos (Reinbold et al., 2022). Despite signs of change post-COVID-19, visual information primarily revolves around health and safety, with workers expressing a need for more diverse information (Tezel and Aziz, 2017). Complex instructions hinder operational performance, and prevalent information management approaches fail to align with individual workers' dynamic task performance needs (Tezel et al., 2016; Grau et al., 2019), since VM is mainly intended for larger groups of people (Tezel et al. 2015). Furthermore, current VM developments remain single-case solutions that hinder knowledge transfer across stakeholders. VM solutions are often manual and located not in close range of the installation area, leading to improvisational practices (Tezel and Aziz, 2017). VM tools, like visual performance boards, remain operation-specific (Reinbold et al., 2022). These developments rely further on trial-and-error efforts.

More systematic and strategic development efforts are essential for developing DVM solutions that provide timely and relevant information for crews and workers. However, the increasing adoption of (manual) VM solutions and the general trend towards digitalization suggest a potential shift in attitudes towards (D)VM developments. These approaches prioritize incorporating workers' perspectives and enhancing process transparency on an individual level through visual tools. To achieve this effectively, there is a need for more integrated and standardized VM developments that facilitate comprehensive knowledge sharing and systematic deployment.

10.2 Objective and Research Approach

The baseline understanding of this research is that construction actors encounter challenges on site because of inaccessible and non-transparent task-related information as well as fragmented communication and collaboration. These challenges are counteracted by workers utilizing decentralized practices and self-guided decision-making (Görsch et al. 2023), which disturbs their installation work (Görsch et al. 2024). The purpose of this work package was to enhance information distribution and transparency on-site to support workers' and crews' situational awareness (Endsley, 1995; Salmon et al. 2009), thus improving their decision-making and overall task performance (Koskela et al., 2018; Reinbold et al., 2020).

A DVM system should provide workers with personalized and essential process and task information. Developing and implementing such a DVM system requires specific prerequisites and requirements. To identify them, various investigations were conducted. First, previous studies on improving on-site actors' task performance, situation awareness, and decision-making through digitalization and VM initiatives in design and construction were analyzed. Additionally, interviews with drywall installers were conducted to study their work demands and challenges, providing a practical basis for developing DVM requirements for construction workers and crews. Next, literature review and interview findings were discussed, validated, refined, and structured in workshops with the research consortium, including VTT, Aalto University, University of Huddersfield, and University of Nottingham. This collaborative discussion continued during research project meetings with industry partners, where practical perspectives were considered, and the feasibility and validity of the DVM requirements were elaborated upon. Finally, interviews were conducted with construction workers to gather their perspectives on the DVM developments driven by VTT, shedding light on the end-users viewpoints.

10.3 Findings from Literature

To address the objectives of this work package, the literature review focused on research studies centered around visual management, workers' task performance, decentralized decision-making, and situation awareness. A selection of 23 papers (Abou-ibrahim and Lappalainen 2022; Barber et al. 1999; Ben-Alon et al. 2014; Brandalise et al. 2022; Conte et al. 2022; Endsley 1995; Galsworth 1997; Görsch et al. 2020, 2023, 2024; Greif 1991; Koskela 2000; Lehtovaara et al. 2022; Loosemore 2014; Pedó et al. 2022; Reinbold et al. 2022; Riekki et al. 2023; Salmon et al. 2009; Seppänen 2022; Seppänen and Görsch 2022; Tang et al. 2014; Tezel et al. 2015; Tezel and Aziz 2017) from the Scopus database was analyzed to identify requirements for Digital Visual Management (DVM) aimed at enhancing information distribution and transparency on-site, supporting workers' and crews' situational awareness, and improving their decision-making and overall task performance.

Requirements and prerequisites for DVM were collected and clustered using Miro. The analysis included 188 requirements and prerequisites, many of which were found to be similar across different studies or worded differently. Therefore, the researchers clustered similar items together. This process resulted in eleven requirements, which are described in Table 5.

Table 5. Requirements based on literature review

DVM Requirements	Description
(1) Simplicity	Simplicity means that information and visuals are presented clearly and concisely. Conciseness ensures that information is stripped of unnecessary complexity and jargon to focus on the essential details for effective task performance. Clarity is of paramount importance so that workers can understand their tasks quickly and without confusion or misinterpretation.
(2) Standardization	This requirement stresses the importance of clearly defining and organizing tasks and information to ensure smooth coordination of task performance. Information should be structured to seamlessly integrate into workflows, covering past, present, and future data. It must be measurable using simple metrics for real-time monitoring and communication across planning levels. Standardization should support workers' self-management while providing centralized coordination and visualizing project milestones. Predictive insights from tracking progress and historical data should assist decision-making, enhancing workers' understanding of project objectives and dependencies.
(3) Availability	Availability of information and visual representations describes the presence of task, progress, and situational data to guide and streamline individual task performance. Information should be pervasive and integrated into workspaces, emphasizing the situational context (see 11) to ensure it supports efficient and effective task performance.
(4) Accessibility	Accessibility describes personalized, situational, easy (at a glance) and secure access to data without personal data being leaked.
(5) Flexibility	Flexibility describes the ability of employees to adapt visual information to their convenience and needs across different devices (smartphones, tablets, screens, screens integrated to process elements, such es elevators, equipment, tools, etc.). This ensures that employees have access to the relevant information and can adapt to unforeseen conditions and situations on site.
(6) Traceability	With the introduction of digital technologies, data traceability becomes crucial for the DVM. It encompasses the way in which data is collected and recorded, stored, linked and analyzed to ensure interoperability between different levels and formats of information.
(7) Reliability	Reliability describes the provision of correct, up-to-date, and trustworthy information, reducing process variability and guaranteeing the quality and resilience of provided information over time, while information needs and task conditions dynamically change.
(8) Relevancy	Relevancy describes the provision of visual information, which is relevant to employees' workflows and performance, based on their specific information needs. By identifying these needs in advance or on-demand information retrieval, it can be ensured that workers have access to the necessary information when needed.
(9) Immediacy	This requirement emphasizes the instant provision of information to support situation awareness, distinguishing immediacy by both time and space. It ensures that information is readily available and visualized at the direct spatial interface between workers and process elements, such as equipment, devices, and screens in direct vicinity. Real-time data collection, analysis, and distribution enable immediate access to information, allowing workers to retrieve it at a glance and ensure straightforward, readily available data.

(10) Goal-orientation	Goal-orientation describes the purposeful collection, storing, structuring, analysis, and distribution of information in a visual format by focusing on aligning the production system specifics with project objectives and operations flow activities. It aims to enhance the quality of on-site collaboration and installation work by assisting and reducing workers' planning efforts to minimize waste and disturbances.
(11) Situational Context	The situational context describes two aspects. First, the need to present visual information within the environmental context so that workers can perceive and understand information more quickly through contextualization. Secondly, it describes how visual information should be targeted to individual workers, crews or trade levels that differ in their information needs to enable efficient and sufficient communication and collaboration between different actors.

10.4 Worker Interviews

The aim of the interviews (the detailed results can be found in TF.1 Workers' requirements and the related conference paper produced in the ACTOR project (Liinasuo, Salonen & Görsch, 2023) and actor-level decisions) was to find out what the work demands are at the construction site, to be met by DVM. Eleven drywall installers were interviewed, and based on the interviews, the demands were identified. Firstly, the demands were classified into the categories of dynamics, uncertainty, and complexity-related demands. Secondly, the classification was performed according to the primary party responsible for the demand. In the latter categorization, the categories are (i) worker, (ii) superior at the site, and (iii) construction project planner-related demands.

The demands that the worker is primarily responsible for are easy to meet (e.g., the demands of being meticulous and proactive in work). The demands that the superior is mainly responsible for can be rather easily tackled as well – and if the worker does not know how to deal with them, (s)he can always ask the superior (e.g., when the location of work tools or work material is not known). However, the demands that those outside the construction site are responsible for, that is, construction project planners and designers, are the most challenging ones (pressing timetable-related issues and various problems with construction drawings and floor plans).

The demands can be scrutinized from the perspective of worker requirements. For instance, the demand for working without a common language can be translated into the following requirement: The worker must also collaborate with such workers with whom (s)he has no common language. Based on this approach, 14 such requirements can be formed in which workers' decisions are involved (e.g., if the site is messy, the worker is required to decide whether to start cleaning the site or working despite the mess), and 13 such requirements in which worker has nothing to decide (e.g., if the worker is unaware of the next task, the worker is required to clarify what the task is, usually by asking it from the superior).

Many of the demands that the worker or the superior is primarily responsible for can be mitigated by DVM, i.e., the management of work with visual information. It is one way to provide workers with task-related, up-to-date information. There are various possibilities to support the workers with visual information using only figures and symbols with no letters and numbers. The benefit of such visual information in construction sites with workers from many countries is that the information is available for everybody at the site, irrespective of the lack of a common language or even alphabet. Examples of visual information without words and numbers are a 3D model of the finished space, a tool for flexible work sharing, a navigator for a lost tool or material, an updated document deliverer, and a shared drawing to be used among the workers. Visual information with also words and numbers allows an efficient way of organizing work among workers with, for example, an info screen.

10.5 Refinement of Requirements through VM-Expert Workshops

The expert workshops served two main purposes. Firstly, they facilitated discussion and clarification of identified descriptions, prerequisites, and requirements of DVM and related fields. This process also validated findings within the broader context of VM in manufacturing and production. Secondly, the workshops explored strategies and approaches for integrating and structuring all the findings. The aim was to enhance workers' and crews' situational awareness, enabling better decision-making and reducing disturbances to their installation work.

Based on three VM-expert workshops, an understanding of DVM for construction professionals was developed as follows: DVM aims to amplify process transparency for individual workers and crews by leveraging digital methods for gathering, storing, linking, analyzing, distributing information in visual formats to enhance on-site communication and collaboration. At its core, this approach prioritizes a human-centric situational understanding to enhance decision-making during task performance, resulting in less disturbed and more ergonomic workflows for construction actors. To realize this vision, Endsley's (1995) situation awareness model and Salmon et al.'s (2009) distributed situation awareness model serve as foundational frameworks. These models offer insights into establishing situational awareness through visual and digital means. Thus, the overarching goal is therefore to perceive, comprehend, and project information (Endsley 1995) in a timely and digital manner to meet the different information needs of individuals and crews (Salmon et al. 2009). To better understand the prerequisites and requirements for developing a DVM system, the findings from the literature research and expert workshops were categorized according to the processing and usage of data: collecting, storing, structuring, analysis, visualization, distribution, interacting, and transforming. All requirements are presented in Figure 25.

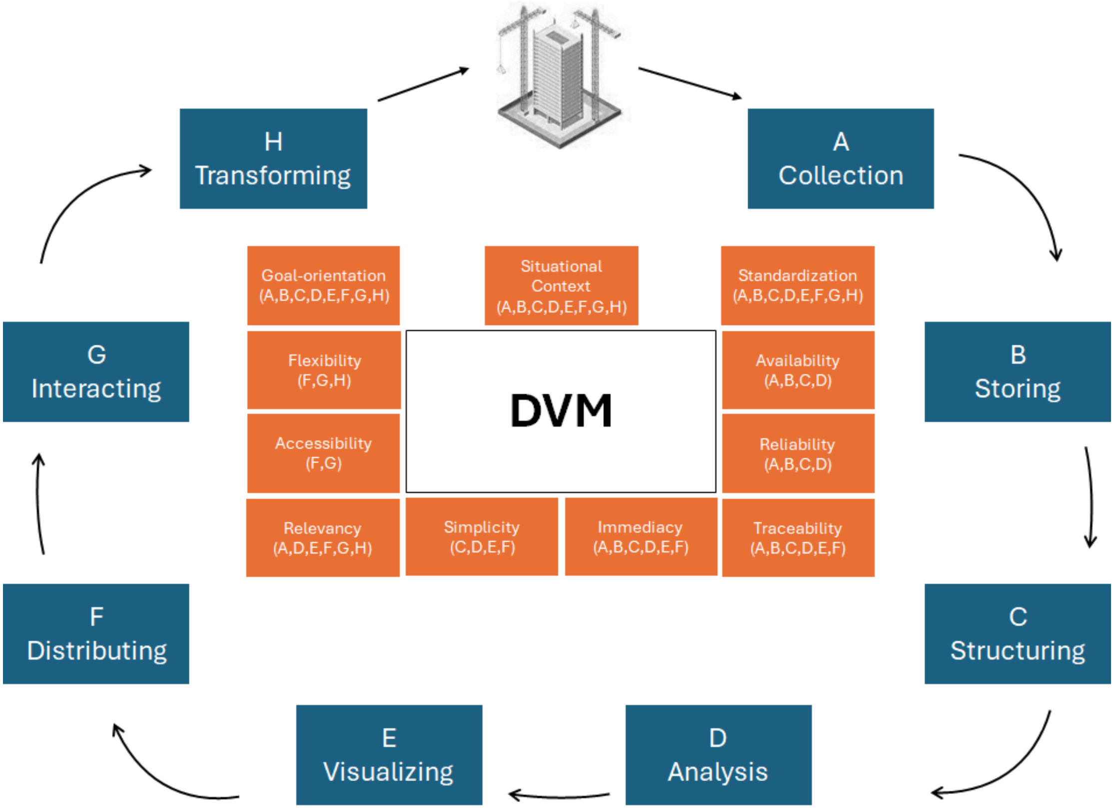


Figure 25: Digital Visual Management system requirements

10.6 Info Screen

Based on the defined requirements, a solution, an info screen, was developed to support workers, informing workers visually about the situation at the construction site. Info screens are large touch displays placed in central locations of the construction site. Shared information can be personalized or work-role-specific. Personalization can be done, e.g., based on an indoor positioning system to identify the worker role or by using face recognition. The same information can also be shown on a hand-held device (Figure 26).

Figure 26. Info Screen being used in a laptop with a touch screen

The Info Screen application (Kiviniemi et al., 2024; this conference paper was produced in the ACTOR project) was programmed using Client-Server web technologies. The visible Client side runs in a standard web browser, and it was programmed using the JavaScript React framework. The server consists of a Java SpringBoot application server and PostgreSQL database, providing the construction project with real-time status. The application server also implements API for receiving data from external applications, such as ActorLeap below. The main visual component, the 3D BIM view, uses the Trimble Connect JavaScript API to show the 3D model stored in the Trimble Connect service.

All the shared information in the Info Screen application is linked to the BIM model, and the user can easily access it by manipulating the 3D model with touch gestures.

<u>Features of Info Screen</u>

The Info screen device helps the workers to:

- find out the next proposed and available sub-activity and work location,
- find locations of material storages and equipment,
- and be aware of no-go areas, alarms, and other exceptional conditions.
- Also, general daily information can be presented.

The images below show the Info Screen user interface. **Error! Reference source not found.** shows anInfo Screen view for a selected floor with a visualization of the work status of the drywall installer. Proposed task is shown in lime colour, another available task in amber colour, another team working nearby in light blue colour, and just finished work location in green colour.

Figure 27. The Info Screen application UI

Figure 28 shows an example of different types of location-based information, including the locations of material storages and equipment and restricted areas.

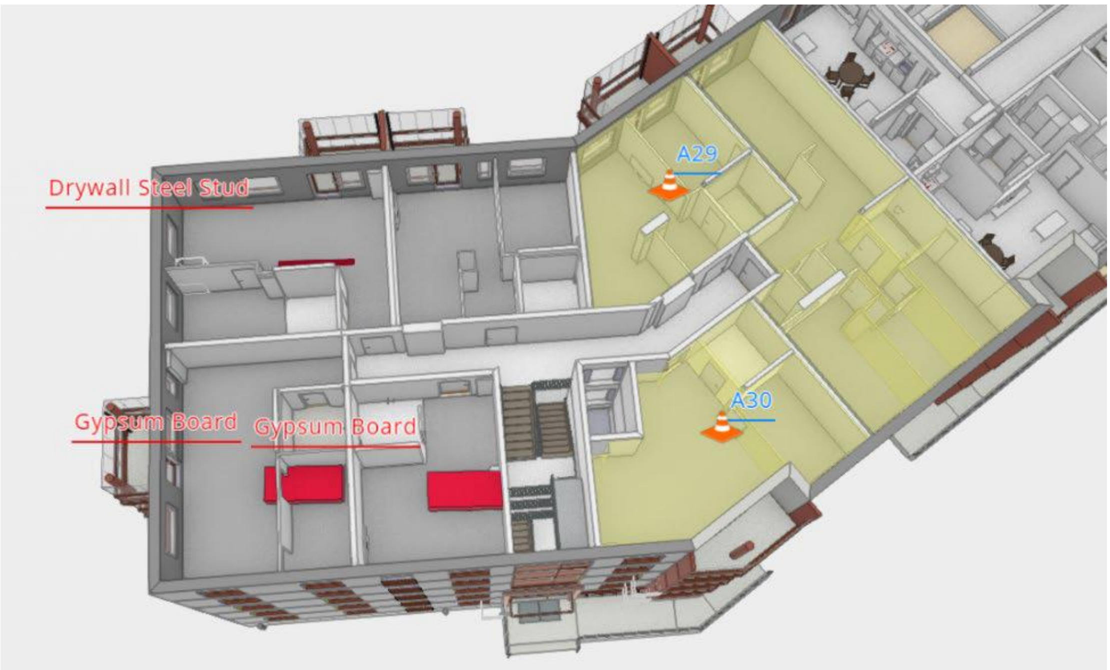


Figure 28. The Info Screen application UI. Examples of visualization of restricted areas and material storages

10.6.1 End-user testing

End-user interviews; background

Construction workers tested the potentiality of Info Screen for future use. Testing was performed at the office next to a construction site. One researcher presented the user interface of Info Screen, conducted the interview, and wrote down the responses. The application provided the following information, located on a three-dimensional model of a floor:

- workspace status (ready; prohibited area; don't go, others work here; recommended for you to work here; a possible working location for you; others work here and also you can come)
- work status for the drywall installer (one board is installed; two boards are installed)
- location of the work material and work tools on the floor
- design documents
- general-level information (free text such as announcements).

Method

Four carpenters also performing drywall installation were interviewed about Info Screen. Their average age was 48,5 years (min 39 years, max 63 years). Work experience was about 30 years on average (min 20 years, max 47 years). Each interview lasted for one hour. During the interview, all main functionalities from the user perspective were presented by guiding the tester to use the device, and testers' opinions were asked about the functionalities during the usage of the device.

Results and discussion

Most testers liked Info Screen and used it easily. All information provided by the screen was found necessary. The oldest tester, unfamiliar with digital technology, had some problems. Still, all testers could rotate and move the 3D model and handle the device well. They all rated the information quality as good (4 on a scale of 1-5, with 1 as poor and 5 as excellent). Most rated the device usability as good, too (4, average 3.25; same scale as above). One said that the device was easy to learn and suggested having short instructions next to it for new users.

The device's information can be displayed on a big screen at the site and/or as a phone app Two testers preferred both options; one liked the big screen more, and one only wanted the phone app. One commented that even if a large screen that everyone can view could be good, it is not nice to look for information when there are other workers rushing behind your back.

The testers were inquired about how much the device could recognize the user's identity and provide customized information accordingly. The options were revealing the individual or the work role. Contrary to expectations, three testers did not object to the device's capability of identifying the worker personally when standing nearby. One of them remarked that the site was already equipped with cameras, and the work pace was so fast that there was no opportunity to do anything else but concentrate on the task. One tester preferred that the device only recognized the person's work role and language preference.

The information about the location of work materials and tools was considered clear and useful. The most useful information was the statuses of the work locations (two responses), easy access to design documents (one response), or both statuses and documents (one response).

The statuses of the work locations elicited the most discussion. Among the responses, only the symbols indicating a completed and prohibited areas were unanimously clear; the symbol indicating a potential work location was the most frequently criticized as confusing. Two testers mentioned that the status information about the work phase of the drywall installation was relevant to the electrician, plumber, and painter; one noted that the information about the presence of other workers in the area would be helpful so that the drywall installer could go there when they were finishing their work. This is consistent with the construction work requiring coordination and collaboration among various work roles.

10.7 ActorLeap

ActorLeap is a solution that provides visual support to workers in performing several types of tasks. In this project, the information is mainly to be used for drywall installation, but many functionalities could also be used in other tasks.

The ActorLeap application was developed for the Magic Leap 2 augmented reality headset using the Unity Editor. The application renders a building's 3D model, and building services engineering 3D models around the user. The building’s 3D model is rendered transparent, allowing the user to see the real building while being able to interact with the augmented building using the glasses and controller. The initial positioning of the glasses on-site is established with marker tracking. The markers are placed identically in the 3D model and the physical site. After the initial pose is initialized, the glasses’ own tracking system is used for tracking. Figures 29-33 show the user’s view through the AR glasses and some use cases. Figure 34 shows a researcher using the system.

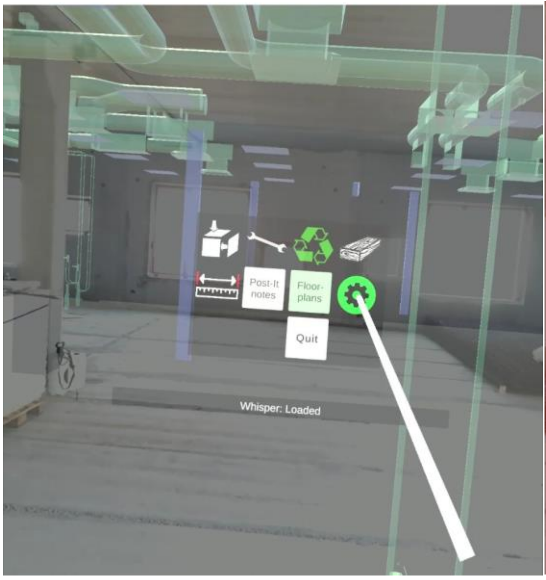


Figure 29. Main menu in the augmented reality

Figure 30. Virtual drywall, floorplan and building services engineering 3D models perceived though AR glasses, as if part of the real world

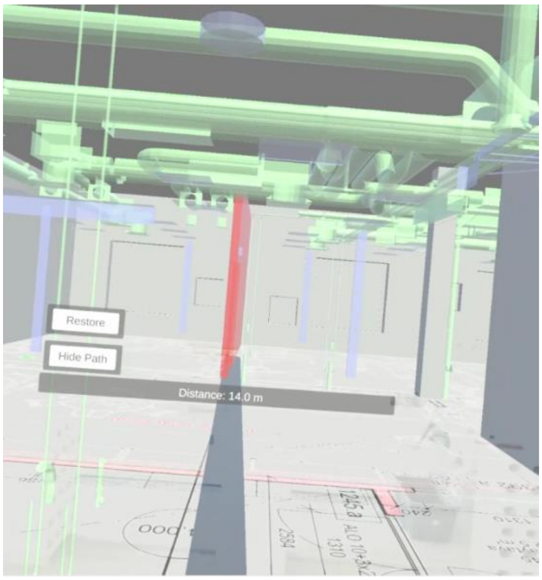


Figure 31. Path (blue line) and distance to the next drywall, perceived through the AR glasses

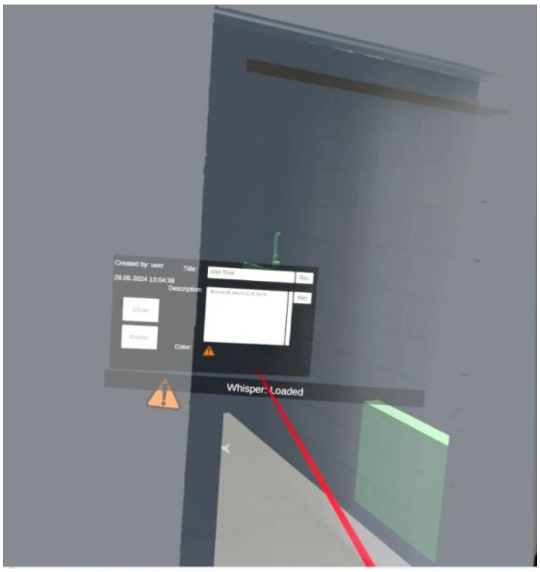


Figure 32.2 Virtual ost-it note in the menu of the AR system and an augmented post-it note icon on the wall

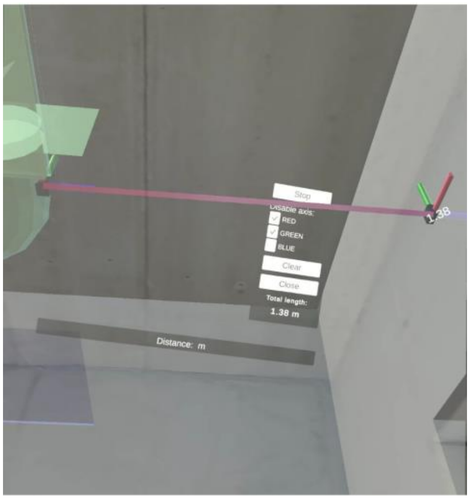


Figure 333. Measuring tool perceived with the AR glasses

Figure 34. User wearing Magic Leap 2 glasses, with a remote control, at a construction site

ActorLeap provides the user with the path (Figure 31) to the next drywall, tools, recycling positions and materials. The path is calculated using UnityEngine.AIModule's NavMesh component.

The measuring tool (Figure 33) is used with the controller. The user first points to the starting position. Then, the user can click as many points as he/she likes and get the measurement for each line, as well as the total distance. The user can freeze any axis to be able to draw straight measuring lines in the desired direction. Clicking positions are chosen based on ray tracing, which requires some 3D content to be present at each position.

A floorplan (visible in Figures 30 and 31) is generated from a PDF file, which is first converted to SVG format. Then, using the Inkscape tool, several small areas are cropped to create PNG images with higher resolution. These PNG images are then aligned with the 3D model at the floor level.

For the creation of virtual post-it notes (Figure 32), users also have the option to use speech-to-text functionality in addition to typing with a virtual keyboard. This voice input feature utilizes the whisper.cpp library (https://github.com/ggerganov/whisper.cpp Accessed: 2024-06-05) with the tiny English language model.

10.7.1 End-user testing

End-user interviews; background

The potentiality of ActorLeap was evaluated among construction workers. The assessment took place at the construction site, within the context of actual work conditions. The content provided by ActorLeap was customized to align with the real tasks of the workers, ensuring the information was applicable and, in theory, could be utilized on-site. Due to the constraints of the testing timeframe, certain simplifications were necessary. Consequently, the location of the materials to be located using the device was proximate, and the materials to be identified were conspicuously placed in the center of the hall where the testing took place. Nonetheless, this did not hinder the evaluation of the device's utility and effectiveness, specifically the precision and productivity of the guidance functionality in directing to the correct location.

One researcher facilitated the participant's use of ActorLeap interface by instructing on its operation and by offering assistance with any issues encountered. This was possible because the researcher had access to the same augmented-reality display as the testing participant wearing the augmented-reality glasses. Another researcher was responsible for conducting interviews and documenting responses. No recording was taken to maintain participant confidentiality. The following functionalities were tested:

- Guidance to the location of the next task (similar with the guidance to a needed tool and guidance to the litter bin): the user chooses the target and the path to the target is shown in augmented reality in blue color
- Measuring: user measures distance between two or more points; as the needed measurement is in one direction, the two other directions can be frozen so that the line is guaranteed to be straight
- Post-it Notes: it is possible to write (by writing with a virtual keyboard or by talking), read, and delete Post-it Notes that they appear to be attached to the real world; in the test, two ready-made notes were used (one warning about wet floor and another warning about a hole in the floor) so that the user only chose the option of viewing the notes; the possibility of creating a Post-it Note was not offered due to time limitations
- Floorplan: when choosing the option Floorplan, the floorplan with measurements appears on the floor, coinciding with the real locations on the floor (in the test, the real floorplan of the location in question was used).

Method

Four carpenters, also performing drywall installation, tested ActorLeap. Their average age was 43,5 years (min 39 years, max 53 years). Work experience was about 22,5 years on average (min 15 years, max 35 years).

When asked before the test whether virtual reality (VR) or augmented reality (AR) was familiar, one tester said that VR was familiar to him as he had played with VR glasses. For others, the meaning of the terms was unclear, even if one had tried the AR/VR glasses and another had played a mobile game with them.

Before the testing, the glasses were put on to verify that the nose piece was appropriate. If it was not, the nose piece was changed to a better one to enable a good fit.

Each test lasted one hour. During the test, one functionality was first used with the researcher's support, and thereafter, another researcher asked questions about it. Then another functionality was tested, etc., until all functionalities were tested.

Results and discussion

All testers were positive about ActorLeap and used it rather fluently. It should be remembered, though, that the researcher was ready to support the tester with the user interface as much as needed. The tasks were quickly performed from the perspective of controlling the device; that also depended on the small number of controlling actions needed to perform the tasks and evaluate the related functionalities. Also, the familiarity of virtual reality glasses in the gaming context has probably influenced the ease of use to at least some of the testers.

Testers were asked to grade the usefulness of ActorLeap in the scale familiar from the school (4=fail, 10=A). The testers rated usefulness with an average of 9.6, which is a very high rating (two ratings of 9 and two with of 10). The usability of ActorLeap was rated as 7.75 on the average (one rating of 6, two ratings of 8 and one rating of 9).

After the testing, the physical ergonomics of the glasses were inquired. AR glasses were not heavy to wear. Only one tester reported afterward that the glasses were ruby; that was probably because of a wrong nose piece (there were several nose pieces for testing, but the defect was not noticed when the AR glasses were on before the testing). The glasses were not loose either, except for one tester, who said afterward that there was a slight "jump" in the glasses when walking, so probably it was loose after all. As a general comment, one tester said that the sides of the device prevented from seeing to the sides; it would be better to have the device integrated with goggles. Another tester suggested the glasses could be integrated with the visor. All testers preferred a remote control as the means of controlling. Other options were gestures, eye movements, and speech. This opinion may also be due to the lack of experience of these other controlling means. Descriptions of the possibilities for the remote control were the following:

- joystick that could be used without keeping your arm straight (that is a preferable pose when using the remote control in the test)
- device in the wrist with touch control or some other remote control that is attached to the user as pockets are already full
- small remote control hanging on the user's neck
- integration with the mobile phone or some other small device.

Regarding functionalities, <u>the guide to the next workspace</u> was found to be easy to understand and use. One tester commented that the system appeared a bit too sensitive, so it was hard to hit the menu in the augmented world. Other testers did not report any difficulty in controlling the device. All commented that the augmented path, a thick blue line on the floor, was a bit hard to see. Regarding the usefulness of this functionality, testers stated that it was good for a new worker and an experienced worker (1 response) when going to the upper floors for the first time (1 response), whenever you need to know where to go next (1 response), and when wanting to know where the wall should be without measuring it (1 response).

Because the blue guiding path was poorly visible and due to a restricted timetable, the testing related to the guidance to a tool and a litter bin was left out. <u>Guide to work material</u> (boards) was tested with three testers, and their comments were positive ("just ok", "the menu was clear", "it was easy to use"). Regarding the usefulness of the functionality, it was stated it could be useful when working for the first day or that when the space is not open, such support would help to find the material behind the walls.

<u>Measuring</u> with the device was found to be easy. Two testers who appeared to be the most skillful in using the device were allowed to lock one dimension in the 3D world so that the line becomes straight in that respect. Only one tester commented that his hand waggles a bit, which makes controlling first harder; others did not report any difficulties. Measurement with ActorLeap was found to be quicker than measuring with measuring tape or laser. It was found beneficial that with the device, there is no need to climb into the site. It was found to be good that the values remain in the augmented world, because the worker then sees, based on measures, where the pipes are running and can take it into account in other tasks. This is a remarkable benefit because, as one tester commented, there is a lot of measuring to do for a worker. Digital measuring was found especially suitable for the drywall installer

because drywall installation involves measuring doors and walls of different heights. Measuring is performed both before and after the work, increasing the need for a handy device.

It was easy to control the device when looking at the Post-it Notes in the augmented world. Testers were asked whether they would create a new note by talking or writing; two preferred talking and two preferred writing (but without any experience or knowledge of how that would be performed with the device). All found this functionality useful. The comments were mostly about the content of the notes and not about the functionality; one stated that the announcement about a space you cannot enter is good to have, and another noted that it would be good to know when the floor will be worked with because workers do not know about each other's work, you must always ask. One tester just commented that it is good to become aware that the floor is wet or has a hole in the floor. One tester invented that the notes could also be used to inform that some specific worker starts in this location at some specific time.

All found floorplan, made visible in the augmented world, useful, and all but one also easy to use. The one with the opposite opinion reflected that it probably becomes easy when you look at the floorplan for some time. There was furniture included and it is not part of the tester's work, which may have affected the opinion. On the other hand, the tester considered that the appearance of furniture supports understanding whether something could be removed. This functionality raised positive comments the most. One tester said he would need this every day (this tester has various tasks such as installing drywalls, making the concrete floors, battening, etc.). One commented that the augmented floor plan would be useful. Presently, the tester looks at the floorplan via mobile phone. Another stated that the constant visibility of the floorplan would support working as you do not have to look constantly at the mobile phone and zoom in there. It was also commented that when a floorplan is easily visible, one can leave tools and materials at the site without unintentionally preventing others from doing their work. The fact the floorplan is instantly in the correct scale makes it better than a floorplan on paper (or mobile phone), said one tester.

There were some suggestions for amendments and additional functionalities. One tester stated that it would be good to know whether there will be something inside the wall; it affects whether the drywall installer can work with both walls immediately or has to wait. Another tester commented that errors made during construction work could be visible - the task of that tester is to correct errors, and the tester considered it would facilitate work to see errors immediately through the glasses. Furthermore, some additional physical light would be needed when looking through the AR glasses in dark locations as presently, the dark space appears white, hiding obstacles that are poorly visible in the dark. One tester invented that it could be possible to tell the application that the person leaves this tool here and the other person could find it then. This functionality, however, would not be needed if the workspace were monitored and a picture of the site is constantly updated. Then, ActorLeap would only tell the path to the tool. Finally, the possibility of measuring square meters was also wished for.

11 Summary and conclusions

The aim of ACTOR project was to automate the coordination of construction actors with a digital situation picture. In the project, we defined the workflow required for automation. Technical tests were done related to each part of the workflow, including automatic data collection, data storage in a common data lake, data analysis with computer vision, agent-based simulation models and modeling worker decision making, and finally, digital visual management to communicate the results to the workers. The developed knowledge can be used by companies to develop commercial solutions. The consortium company partners Trimble, Flow Technologies, CarinaFour and Kone have already used some of the research results in their development.

Many solutions developed and tested in the research projects are leading to publications in scientific conferences and journals. The ontology extension work has resulted in one journal paper and two conference papers during the project and one additional journal paper has been submitted for review, and one is in preparation. The computer vision tasks will result in at least one journal paper. Automated data collection includes plans for three journal papers. Digital visual management requirements and solutions have been published in two conference papers. The academic output of the project will be 10-15 publications, which is in line with the planned academic deliverables.

However, although all parts of the workflow were tested separately, there were challenges in testing the combination. Developing computer vision is time-consuming, and thus, the project scope was limited to drywall installation. The drywall challenge was solved well, but the other data streams, particularly indoor positioning data, were not available for drywall due to a lack of informed consent from drywall workers. Indoor positioning data is critical input information for agent-based simulation, so the simulation model had to be done with another task: furniture installation. This broke the chain from data collection to visual management because all other tasks were based on drywall installation. This limitation should be addressed in future research by ensuring the consent for indoor positioning before selecting task types for computer vision analysis and visual management.

The results presented in this final report and scientific publications offer useful starting points for companies developing their systems. Results related to automatic data collection and implementation of the data lake and digital visual management systems are likely to be useful to practitioners. The use of simulation models and digital twins of processes will likely require additional research work before companies are able to use the results. The PI has received research funding from the Research Council of Finland, which allows continuing work on this track.

References

Abou-Ibrahim, H. and Lappalainen, E. (2022) Digital Twin of a Design Process: An Exploratory Study. *Annual Conference of the International Group for Lean Construction*, pp. 1198–1209.

Abou-Ibrahim, H., Seppänen, O., Al Barazi, A., & Görsch, C. et al. (accepted) Automatic Operational Decision-Making on Construction Sites using Choosing by Advantages and Situational Awareness. Accepted for presentation and publication in EG-ICE 2024 workshop in Vigo, Spain.

Barber, P., Tomkins, C. and Graves, A. (1999) Decentralised site management - A case study, *International Journal of Project Management*, 17(2), pp. 113–120.

Ben-Alon, L., Sacks, R. and Grobman, Y. J. (2014) Similarities and Differences between Humans' and Social Insects' Building Processes and Building Behaviors. *Construction Research Congress*, pp. 51–60.

Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. *Automation in construction, 114*, 103179.

Brandalise, F. M. P., Formoso, C. T. and Viana, D. D. (2022) Development of a Typology for Understanding Visual Management Concepts and Their Relationships. *Journal of Construction Engineering and Management*, 148(7), pp. 1–14. doi: 10.1061/(asce)co.1943-7862.0002300.

Chauhan, I., & Seppänen, O. (2023). Automatic indoor construction progress monitoring: challenges and solutions. In *EC3 Conference 2023 (Vol. 4). European Council on Computing in Construction.*

Conte, M. *et al.* (2022) Exploring the Use of Digital Visual Management for Last Planner System Implementation. *Annual Conference of International Group of Lean Construction*, pp. 645–656.

Endsley, M. R. (1995) Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), pp. 32–64.

Fathi, H., Dai, F., & Lourakis, M. (2015). Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges. *Advanced Engineering Informatics, 29(2)* ,149-161.

Galsworth, G. (1997) *Visual Systems: Harnessing the Power of the Visual Workplace*. American Management Association.

Görsch, C. *et al.* (2020) Construction Workers' Situational Awareness – An overlooked Perspective. *Annual Conference of the International Group for Lean Construction*. Berkeley, pp. 937–948.

Görsch, C., Al Barazi, A., Seppänen, O., & Abou Ibrahim, H. (2022). Uncovering and Visualizing Work Process Interruptions Through Quantitative Workflow Analysis. *Proc. 30th Annual Conference of the International Group for Lean Construction (IGLC)*. https://doi.org/10.24928/2022/0116

Görsch, C., Seppänen, O., Peltokorpi, A., & Lavikka, R. (2024). Task planning and control in construction: revealing workers as early and late planners. *Construction Management and Economics*, 42(5), 431–450. https://doi.org/10.1080/01446193.2023.2270080.

Görsch, C., Seppänen, O., Peltokorpi, A., & Lavikka, R. (2024). Unlocking Productivity: Revealing Waste and Hidden Disturbances Impacting MEP Workers. *Journal of Construction Engineering and Management*. DOI: 10.1061/JCEMD4/COENG-14204

Grau, D., Abbaszadegan, A., & Assanair, R. (2019)'Process versus operations workflow - Making the case for continuous monitoring of construction operations. *Annual Conference of the International Group for Lean Construction*, pp. 563–572.

Greif, M. (1991) *The Visual Factory - Building Participation through Shared Information*. Portland, Oregon: Productivity Press.

Hamledari, H., McCabe, B., & Davari, S. (2017). Automated computer vision-based detection of components of under-construction indoor partitions. *Automation in Construction, 74*, 78-94.

Kiviniemi, M., Charlesraj, V. P. C., Liinasuo, M., Siltanen, P., Rainio, K. (2024) Digital Info Screen – A Visual Management Tool for Construction Workers. *ISARC 2024 Lille, France, June 3-5, 2024*.

Koskela, L. (2000) *An exploration towards a production theory and its application to construction*. Espoo: Finland: VTT Technical Research Centre of Finland.

Koskela, L., Tezel, A. and Tzortzopoulos, P. (2018) Why Visual Management?. *Annual Conference of the International Group for Lean Construction*, pp. 250–260.

Lehtovaara, J., Seppänen, O. and Peltokorpi, A. (2022) Improving construction management with decentralised production planning and control : Exploring the production crew and manager perspectives through a multi-method approach. *Construction Management and Economics*. Routledge, pp. 1–24.

Liinasuo, M., Salonen, T.-T., Görsch, C. (2023) Drywall Installers' Work Demands – Tackling between Normal Duties and Absurd Challenges. *Conference ECCE 2023 (European Conference of Cognitive Ergonomics)*, Swansea, Wales, UK, on 20-22 September, https://dl.acm.org/doi/abs/10.1145/3605655.3605681

Loosemore, M. (2014) Improving construction productivity: a subcontractor's perspective. *Engineering, Construction and Architectural Management*, 21(3), pp. 245–260.

Masood, M. K., Aikala, A., Seppänen, O., & Singh, V. (2020). Multi-Building Extraction and Alignment for As-Built Point Clouds: A Case Study With Crane Cameras. *Frontiers in Built Environment, 6*, 581295.

Mesa: Agent-based modeling in Python — Mesa .1 documentation. (n.d.). Retrieved May 13, 2024, from https://mesa.readthedocs.io/en/stable/

Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Bai, Y., & Behnood, A. (2022). 3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review. *Advanced Engineering Informatics, 51*, 101501.

Noccela Oy. (n.d.). *Noccela*. Retrieved September 18, 2023, from https://noccela.com/

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering, 22(10)*, 1345-1359.

Paneru, S., & Jeelani, I. (2021). Computer vision applications in construction: Current state, opportunities & challenges. *Automation in Construction, 132*, 103940.

Pedó, B. *et al.* (2022). Visual Management Requirements to Support Design Planning and Control within Digital Contexts. *Sustainability*, 14(17).

Rakennustieto. (2023). *Aikataulukirja 2024 (Ratu KI-6036)*. https://kortistot.rakennustieto.fi/kortit/Ratu KI-6036

Reinbold, A., Seppänen, O. and Peltokorpi, A. (2020). The Role of Digitalized Visual Management to Empower Selfmanaged Crews in Construction Projects. *Annual Conference of the International Group for Lean Construction*. doi: 10.24928/2020/0021.

Reinbold, A. *et al.* (2022). Current Challenges in the Adoption of Digital Visual Management at Construction Sites : Exploratory Case Studies. *Sustainability*, 14(21), p. 14395.

Riekki, J. , Rannisto, J. , Lehtovaara, J. , Seppänen, O. & Peltokorpi, A. 2023. Achieving a 4-Hour Takt Time – and Driving Change With It. *Annual Conference of the International Group for Lean Construction*

Roh, S., Aziz, Z., & Pena-Mora, F. (2011). An object-based 3D walk-through model for interior construction progress monitoring. Automation in Construction, 20(1), 66-75.

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S., & Girolami, M. (2020). Construction with digital twin information systems. *Data-Centric Engineering, 1*, e14.

Salmon, P. M. et al. (2009) *Distributed Situation Awareness. 1st edn*. London: CRC Press.

Seppänen, O. (2022) Use of the Digital Situation Picture to decrease Waste in the Design and Construction. in *Lean Construction 4.0 - Driving a Digital Revolution of Production Management in the AEC Industry*.

Tang, P. *et al.* (2014) Workflow stabilization with finegrained work packaging and near real-time progress monitoring. *Annual Conference of the International Group for Lean Construction*, 1(602), pp. 739–750.

Tezel, A. *et al.* (2011) An examination of visual management on Finnish construction sites. *Annual Conference of the International Group for Lean Construction*, pp. 115–124.

Tezel, A. *et al.* (2015) Visual Management in Brazilian Construction Companies : Taxonomy and Guidelines for Implementation. doi: 10.1061/(ASCE)ME.1943-5479.0000354.

Tezel, A., Koskela, L. and Tzortzopoulos, P. (2016). Visual management in production management: a literature synthesis. *Journal of Manufacturing Technology Management, Vol. 27 No. 6*, pp. 766-799. https://doi.org/10.1108/JMTM-08-2015-0071

Tezel, A. and Aziz, Z. (2017). Visual management in highways construction and maintenance in England. *Engineering, Construction and Architectural Management*, 24(3), pp. 486–513.

Tezel, A., Koskela, L. L. and Tzortzopoulos, P. (2010). Visual management in construction: Study report on Brazilian cases. *SCRI Research Report*, (March), p. 36.

Yang, J., Park, M. W., Vela, P. A., & Golparvar-Fard, M. (2015). Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future. *Advanced Engineering Informatics, 29(2*),211-224.

Zhao, J., Seppänen, O., Peltokorpi, A., Badihi, B., & Olivieri, H. (2019). Real-time resource tracking for analyzing value-adding time in construction. *Automation in Construction*, *104*, 52–65. https://doi.org/https://doi.org/10.1016/j.autcon.2019.04.003

Zhao, J., Pikas, E., Seppänen, O., & Peltokorpi, A. (2021). Using real-time indoor resource positioning to track the progress of tasks in construction sites. *Frontiers in Built Environment, 7*, 661166.

Zheng, Y, Törmä, S., & Seppänen, O. (2021). A shared ontology suite for digital construction workflow. *Automation in Construction 132: 103930.*

Zheng, Y., Khalid Masood, M., Seppänen, O., Törmä, S., & Aikala, A. (2023). Ontology-Based Semantic-Construction Image Interpretation. *Buildings, 13(11)*, Article 2812. https://doi.org/10.3390/buildings13112812

Zheng, Y., Seppänen, O., Masood, M.K., Törmä, S. (2024). Ontology-Based Construction Process Library for Process States Inference. In: Skatulla, S., Beushausen, H. (eds) Advances in Information Technology in Civil and Building Engineering. *ICCCBE 2022. Lecture Notes in Civil Engineering, vol 357.* Springer, Cham. https://doi.org/10.1007/978-3-031-35399-4_32