DID-C1 Technical Report

Geomatics Synthesis Project, spring 2016

F. J. Bot, H. H. Braaksma, R. C. Braggaar, B. R. Ligtvoet, B. R. Staats

TU Delft

in_sight

in_sight

Technical Report

F. J. Bot, H. H. Braaksma, R. C. Braggaar, B. R. Ligtvoet, B. R. Staats

Technical Report
MSc Synthesis Project
Geomatics for the Built Environment
Delft University of Technology
Spring 2016

Supervisor:	Dr. Ir. B. M. Meijers
Course Coordinators:	Dr. Ir. S. C. van der Spek
	Ir. E. Verbree

Contents

PREFACE 7

Abbreviations 7
Glossary 7

1. EXECUTIVE SUMMARY 9

1.1. Introduction 9
1.2. Methodology 10
1.3. Cross Cutting Topics 15
1.4. Conclusions 15
1.5. Recommendations 16

2. INTRODUCTION 17

2.1. Problem Description 17
2.2. Research Subjects 18
2.3. Research Question 20
2.4. Research Scope 22
2.5. Reading Guide 24

3. METHODOLOGY 25

3.1. Rich Picture 25
3.2. Requirement Analysis 25
3.3. Risk Register 28
3.4. Organisation 28
3.5. Research Methodology 30

4. DATA CAPTURE 31

4.1. WiFi systems 31
4.2. AP System 32
4.3. Data Storing 36

5. DATA ANALYSIS 37

5.1. The need for filtering 37
5.2. Visits between SLs in relation to traffic 38
5.3. Filtering methods and grouping 38
5.4. Effects of Filtering and Grouping on Data Flow 43

6. DATA STORAGE 44

6.1. Database Structure 44
6.2. Processing new data 46

7. DATA COMMUNICATION 48

7.1. Boundary Conditions 48
7.2. Dashboard 48
7.3. Progress 51

8. DATA QUALITY 52
8.1. Data Capture 52
8.2. Data Analysis 59
8.3. Data Storage 62
8.4. Communication 63
8.5. Conclusion 63

9. CROSS CUTTING TOPICS 64
9.1. Privacy 64
9.2. Validity and Accuracy 64
9.3. Representativeness 65
9.4. System of APs 66

10. EVALUATION 68
10.1. Occupation 68
10.2. Exploitation 68
10.3. Reliability 68
10.4. User Types 68
10.5. MoSCoW Rules 69
10.6. Results: Knowledge Gained 70
10.7. Hypothesis Evaluation 70

11. CONCLUSION 72

12. RECOMMENDATIONS 73
12.1. General Remarks for the Eduroam System 73
12.2. Recommendations for the in_sight System 73
12.3. Scenarios for Future Application 74

13. REFERENCES 78

Preface

The final project for the first year of the Master of Science Geomatics for the Built Environment at the Delft University of Technology is the Synthesis Project. The project provides students with an opportunity to participate in a real-world assignment containing the full Geomatics process of gathering, storing, processing and delivering spatial data. In the eight weeks of the Synthesis Project all the technical skills gained in the first year of the MSc Geomatics will be used. This report contains the work of project in_sight, which focuses on the occupation and exploitation of the buildings of the TU Delft campus using WiFi monitoring.

Abbreviations

AP	WiFi Access Point
BAG	Register of Addresses and Buildings (Basisregistraties Adressen & Gebouwen)
FMRE	TU Delft Facility Management & Real Estate
ICT	Information & Communications Technology
SL	Spatial Level
O&S	TU Delft Education & Student Affairs (Onderwijs & Studentenzaken)
TU Delft	Delft University of Technology
UFA	Useful Floor Area
SDSS	Spatial Decision Support System

Glossary

SL0	Campus Level
SL1	Facility Level
SL2	Floor Level
SL3	AP Level
SL4	Room Level
Session	Time in which a user is recorded by a single AP.
Visit	A visit depends on the SL. For SL0 and SL1 a visit is the total consecutive session per person in an educational facility. On SL2 it represents the total time of consecutive sessions spend on the same floor in an educational facility.
Record	A row in the *wifilog* table in database.
Occupation	Number of persons on an SL at a specific date and hour
Exploitation	Relative indication for square meters [m^2] per person [n], according to ideal occupation. The ideal is set at 14.5 m^2/n, which is represented by a percentage of 100%.

1. Executive Summary

This executive summary provides an overview of the work done by project in_sight for the TU Delft MSc Geomatics for the Built Environment. The research subject is the usage of WiFi monitoring data to calculate the occupation of the TU Delft campus in order to determine the exploitation of the educational facilities.

1.1. Introduction

In this project, the TU Delft Facility Management and Real Estate department (FMRE) works together with students of the MSc Geomatics for the built environment, to gain insight into occupation and exploitation of available space, inside buildings of interest. The occupation and exploitation numbers are to be analysed on different Spatial Levels (SLs) and visualised through an online dashboard application. The research thereby answers the following research question:
To what extent can the alignment of occupation and exploitation of educational facilities on different scales be indicated through WiFi monitoring?

The research question entails the following research subjects:
- Occupation, as the number of persons [n] on an SL at a specific date and hour.
- Exploitation, as the square meters per person [m^2/n] on an SL at a specific date and hour.
- Alignment, as the rate to which the actual exploitation confers to the ideal exploitation. This ideal can be defined by the user, and is by default set to the desired value 14.5 m^2/n by FMRE.
- Educational Facilities, as the buildings in scope defined by FMRE.
- Different scales, as the SLs researched. These are campus (SL0), facility (SL1), floor (SL2), Access Point (AP) (SL3) and room (SL4).

The research question is evaluated for each defined SL, according to the following hypothesis:

H1: The alignment of occupation and exploitation of educational facilities **can** be indicated through WiFi monitoring on all SLs with adequate reliability.

H0: The alignment of occupation and exploitation of educational facilities **cannot** be indicated through WiFi monitoring on all SLs with adequate reliability.

The hypotheses are evaluated according to the methodology, as described in paragraph 1.2. The scope of the research can be found in figure 1. The WiFi monitoring data are stored into the *wifilog* database table. After processing the data is divided for each of the defined SLs, after which analysis can take place per SL. The analysis is based raw data as well as aggregated and processed data. After this analysis, the occupation values can be calculated. When combined with UFA data and the ideal value, the exploitation can be derived. The occupation and exploitation are linked to geometry on the dashboard, for output visualisation. The research scope entails a process of linking data and information sources through four enrichment stages: data, information, knowledge and wisdom. The gained knowledge is returned to the client through the visualisation, after which FMRE can combine this output with their available sources to gain wisdom as a basis for decision-making.

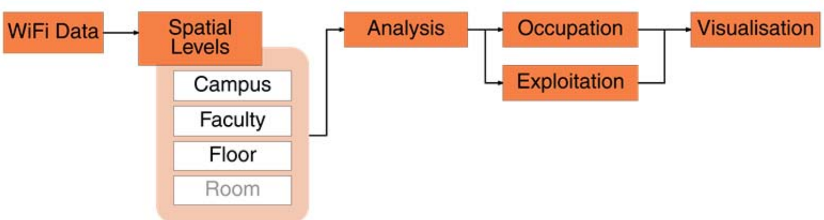


Figure 1: Research Scope.

1.2. Methodology

The research methodology is based on the process for handling geo information of Lemmens (1991, p. 10) (Figure 2), which distinguishes data capture, storage, analysis and communication. As storage of the captured data is in the hands of the TU Delft ICT department, and the storage of the analysed data requires extensive database structures as well, only this partition of data storage is dealt with. The communicated data, as output of the project, can be used by FMRE for management of the physical world situation the process dealt with. The quality of the data is monitored throughout the process by means of performing multiple case studies and validation tests.

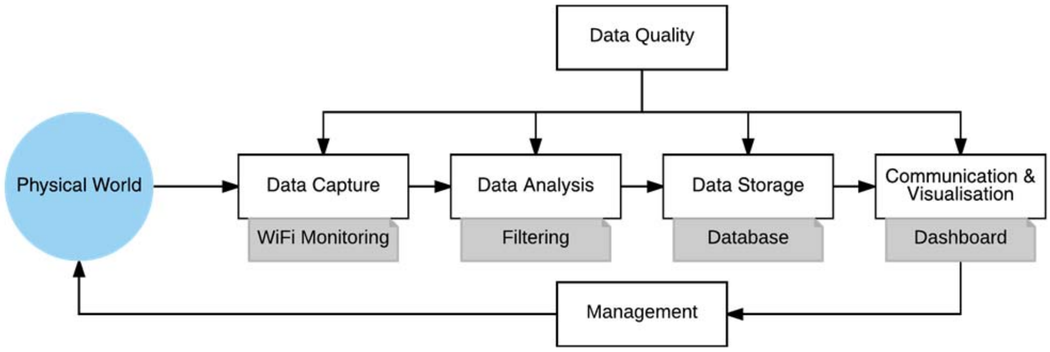


Figure 2: Research Methodology (Adapted from Lemmens, 1991).

1.2.1. Data Capture

The TU Delft AP system, installed for optimal WiFi connectivity, can be used as a tracking system for active WiFi monitoring (Mautz, 2012), in the sense that it stores information on all connection responses executed. The recorded responses are from Eduroam users only, so TU Delft visitors or LAN users are not included into the data. The TU Delft WiFi network is broadcasted through different types of APs, possibly chosen according to building construction and availability of APs on the market. Each AP emits signals according to either the dipole or omnidirectional antenna radiation pattern (Cisco, 2014d).

Data Storage

The recorded responses are stored in a database, referred to as the *wifilog* table. A single entry within this database is referred to as a record, and contains the following information:

- *Username*. A hashed code as a unique identifier for each Eduroam user.
- *Mac*. A hashed code as a unique identifier for each connected device.
- *Asstime*. The timestamp a specific record is associated with. It is the starting time of a session and contains of a date of format yyyy-mm-dd and a time of format hh:mm:ss.
- *Apname*. A TU Delft specific name for the AP in question of the following format: A-X-Y-ZZZ, where X stands for building number, Y stands for floor or wing in select faculties and Z stands for a specific AP number.
- *Maploc*. A description of the AP location with building number, building name and floor, which may be as specific as “ground floor+” or as unspecific as “2nd, 3rd and 4th floor”. Many of these descriptions contain spelling mistakes and discrepancies.
- *Sesdur*. The duration of the recorded session, on a time interval of around five minutes.
- *SNR*. Signal to noise ratio as a measure of signal quality emitted from the AP.
- *RSSI*. Signal strength that a device receives from the AP in dBm.
- *Importfile*. A specification to which file added the data on which data.

System Limitations

The capturing process is prone to several errors. The APs are set up in such a way that the WiFi coverage inside TU Delft facilities is optimal for data communication. However, the current system of APs makes it difficult to track the location of a user based on the recorded data. The emitted signal penetrates walls and floors, and a user is connected to the AP with the best signal reception, rather than the nearest AP. The users' location is assumed to be that of the connected APs' *maploc* attribute in the *wifilog*. Furthermore, the maximum broadcasting power for each AP is variable. The reason for this is twofold. Firstly, in case of failure of an AP, surrounding APs emit a stronger signal as to bridge the gap. Secondly, as to ensure no overload on an individual AP the AP network balances the load between the individual APs.

1.2.2. Data Analysis

To make the raw session data fit for use, filtering is applied (Lemmens, 2011, p. 32). The cases used to filter the data are as follows: users living on campus that can connect to the Eduroam network from their home address, static devices such as printers and traffic (Figure 3).

Figure 3: Cases to be Filtered Out: Users Living on Campus, Static Devices and Traffic.

The static devices and users living on campus must be identified before their sessions can be removed from the data. The sessions by users who live on campus need to be removed before the data is aggregated into visits, because the sessions are identified by username and a specific AP. The visit creation is needed to identify traffic. After the visits creation, the visits generated by static devices are removed. Traffic is removed before the occupation is calculated. The occupation numbers are combined with the floor area per building to create the exploitation numbers. The percentages in figure 4 show the effect the filtering has on the records in the database.

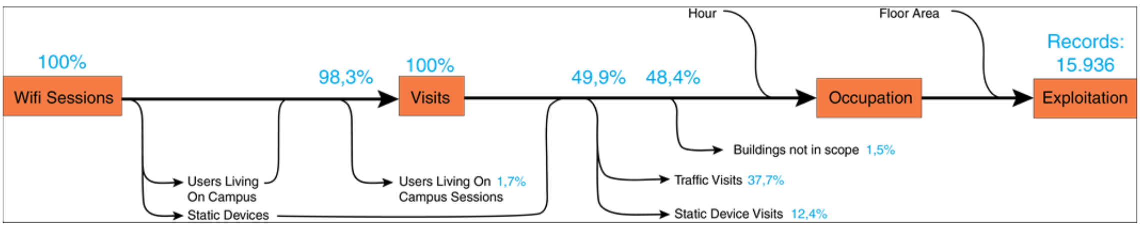


Figure 4: The Filtering Process and its Results.

1.2.3. Data Storage

After the data is processed, it is stored into a set of materialized views, tables and functions in the database (Figure 5). The results of the queries for filtering and aggregating are stored into materialized views. The connected functions ensure that when data is added to the *wifilog*, this new data is processed and appended to tables containing occupation and exploitation columns. In addition to the output tables, additional tables are stored in the database to enable the calculation of exploitation and the visualization. The former are tables containing UFA data and the latter are tables containing different types of geometries for representation on different SLs.

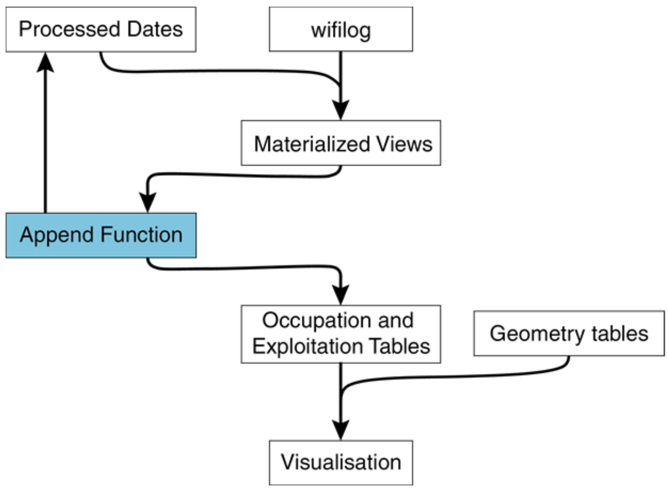


Figure 5: Simplified database structure with materialized views, tables and functions used to process new data.

1.2.4. Data Communication

In order to communicate the output in an intuitive manner, a dashboard web application is created (Figure 6). Several visualisation options are available. The output can be displayed for SL0, SL1 and SL2. SL1 contains the additional option of displaying bar chart geometry (Figure 7), in order to enable the user to directly compare displayed exploitation values to the set ideal. Furthermore, graphs can be displayed containing the retrieved information per day, for either the entire campus or selected facility.

Figure 6: The in_sight occupation and exploitation dashboard showing the building geometry of facility level data. The bar charts show data for the conference center.

The data visualisation can be changed through several parameters. A date between April and June 2016 can be selected. The statistics show the retrieved data per day, while the geometry classification displays hourly data. A custom value can be set for the ideal exploitation value. The dashboard also contains several additional pages, providing information on the project, as well as the information required when publishing personal data, in accordance to the European Data Protection Directive (European Union, 1995).

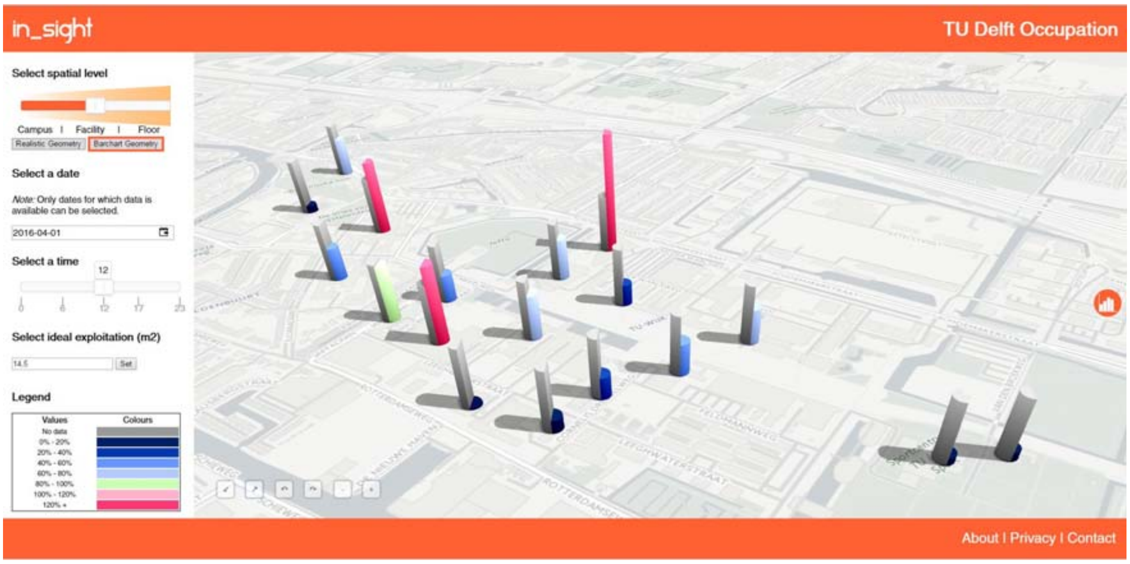


Figure 7: Dashboard Showing Bar Chart Geometry for SL1.

1.2.5. Data Quality

For each step in the methodology, the quality control is performed through theoretical research, multiple case studies and the gathering and analysis of validation data. Each step may be influenced by different aspects.

Data Capture

Four case studies have been performed to define data representativeness on SL2 and three case studies to examine the possibility for a drill down towards SL4. One case study was performed on the top floors of the EWI building, to test AP connectivity across floors. The main goal was to test whether a user would automatically connect to an AP on the floor they are at.

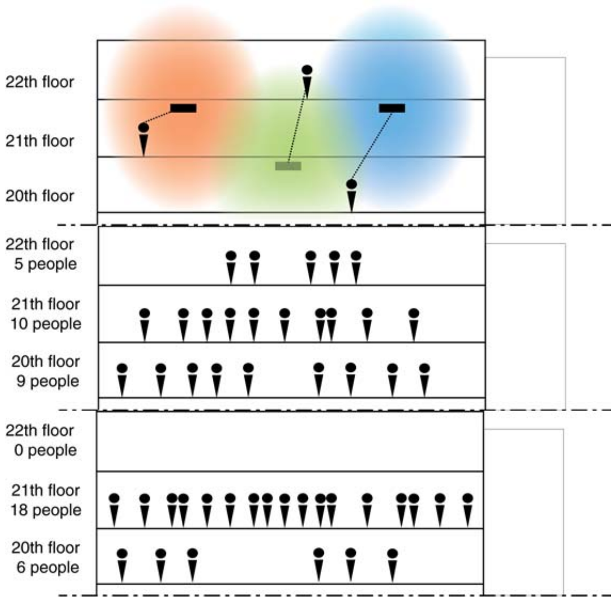


Figure 8: Effects of WiFi System Settings on SL2, Directly Derived from Case Study.

As can be seen in the top of figure 8, Eduroam users are often connected to an AP on a different floor. The WiFi signal emitted appears to propagate further in a vertical than horizontal direction, as users are always connected to an AP on the same side of the building, however often on a different floor. In reality, this may be caused by overlapping signals, competing for emission of the strongest signal. In the data, this results in a misplacement of users based on the *maploc* tags. In figure 8, five people are located on the top floor. However, the processed WiFi system detects them on the AP of the middle floor. The case study implies that a middle floor will be over-represented. The case studies indicate that SL2 data can be derived, however with a low reliability. For SL4 output, the current WiFi system set up does not provide sufficient information.

Data Analysis

To validate the quality of the filtering process, three validation tests have been undertaken. The validation tests are executed at the TU Delft library, a building with a single entrance. An automatic validation system is installed in the building entrance. These counting cameras however have been deemed unviable, as the offset between manual counting and counting camera output is too large. A subsequent validation test, concerning the manual counting session by counting people who are entering or leaving the TU Delft Library per minute, resulted in deviations between 10%-50%. The caused offset is either due to errors in the processed data, or blunders in the manual validation. Therefore, future research should focus on more extensive validation tests.
After finishing the basic filtering process, the quality of the queries is tested through a series of case studies. The quality refers to the extent to which the query covers all aspects to be considered when accurately calculating the occupation and exploitation. Furthermore, the used thresholds are referenced and tested, in order to prevent aggressive filtering.

Data Storage

The quality of data storage is ensured using a crontab. The crontab is connected to a function that ensures the processing of new data added to the *wifilog*. However, when a query is improved, the changes will only apply to new, unprocessed data. A future improvement on stored data would be the refreshing of already processed data after query improvement.

Data Communication

Recurring quality control on communication of results allows a streamlining of the project output to what the client expects. Thus, the client is enabled to adapt the set requirements to insight gained from preliminary results. Active communication between the client and the team has allowed for the production of both reliable and desired results. Outcomes of this type of quality control are the option to input ideal exploitation into the database and the incorporation of bar chart geometry as a means for comparing results.

1.2.6. Management Directions

In the wisdom stage of the research scope, the client can use the gained knowledge for portfolio management and the associated decision-making. Though the retrieval of this wisdom is in the hands of FMRE, some conclusions can be made based on the retrieved data.

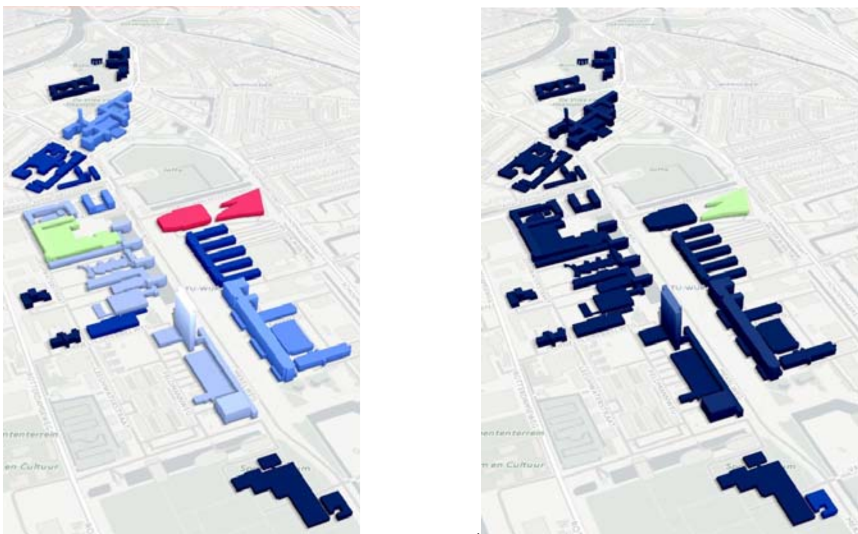


Figure 9: (A) Exploitation at Tuesday 19-04-2016 14:00 and (B) Sunday 24-04-2016 14:00.

Figure 9 depicts two situations of exploitation during a regular lecture week. In general, most facilities return the blue 'calm' classification, in different gradations. Figure 9.A shows ideal exploitation for the faculty of Industrial Design, where other faculties are either rather calm. Both the Aula and Library are greatly over occupied at the displayed time, most likely caused by lunchtime. Figure 9.B shows that opening just the TU Delft library on a regular Sunday afternoon is sufficient for desired use. By exploring more cases like these two examples, FMRE can gain insight into occupation and exploitation of available space through the created the dashboard.

1.3. Cross Cutting Topics

Four cross cutting topics occur throughout the project, which are discussed due to their social and technical relevance.

Privacy

According to the European Data Protection Directive (European Union, 1995), the *wifilog* deals with personal data. Therefore, all data processed should be handled with care, according to the specific directive instructions. All data is gathered from a system that requires subscription to the university the research is conducted for. Therefore, research purposes provide legal ground for data gathering. The published data should be accompanied by the identity of the controller, the purposes of processing, the categories of data concerned, the recipients and the existence of the right of access and the right to rectify ones data. This information is published onto the dashboard, as required.

Validity and Accuracy

Validity is defined as well grounded and justifiable (Merriam-Webster, 2016) and is controlled by ensuring the processing of all available useful information. This is done through the assignment of both superfluous and deficit data. Accuracy is defined as conformant to a certain standard and free of error (Merriam-Webster, 2016), and can be ensured by validating collected measurements to ground truth, as attempted in the validation tests.

Representativeness

Representativeness is partly covered by data accuracy, as it should provide a comparison between registered data and ground truth. However, it takes a broader perspective onto data that cannot be labelled with accuracy. Underrepresentation may be reached through the fact that only users connected to the Eduroam network are recorded. Overrepresentation may be caused by unjust recording of users to an AP *maploc* or unjust aggregation of data.

System of APs

The available system of APs is concluded insufficient for reliable WiFi monitoring at SL2-4. This is due to the specifications of the system, the characteristics of WiFi signal propagation and the types of data recorded in the *wifilog*. Despite the limitations, gathered data for SL0-1 is deemed sufficiently reliable.

1.4. Conclusions

Project in_sight aimed at creating a smart and reusable workflow, which efficiently identifies the occupation and exploitation of the educational facilities at the TU Delft Campus. The primary hypothesis H1 is rejected based on the performed research. It can therefore be concluded that the occupation cannot be determined for all SLs.

With the obtained insight in the research subjects and after answering the hypothesis, the following research question can be answered:

To what extent can the alignment of occupation and exploitation of educational facilities on different scales be indicated through WiFi monitoring?

The occupation and exploitation can be indicated on SL0 and SL1. The occupation and exploitation on SL2 is also taken into account, although the reliability of the outcome is insufficient. As SL2 is of high interest for research into WiFi monitoring, additional research is proposed into gaining reliable output for occupation and exploitation. The online dashboard can provide preliminary insights to FMRE on the alignment of occupation and exploitation on the Campus, Building and Floor level of the educational facilities at the TU Delft.

1.5. Recommendations

To improve the reliability of the resulting occupation and exploitation numbers, several recommendations are appointed in the following three categories: improvements of the Eduroam system, improvements of the in_sight system and future applications.

- Improvements of the Eduroam system:
  - Consistent naming for all the APs names and locations.
  - Store data from the TUvisitor network.
  - Store LAN network data.
  - Perform a more in depth research on the anonymous user account.
- Improvements of the in_sight system:
  - Improve the proposed queries.
  - Process the WiFi data into smaller chunks.
  - When queries are improved, reprocess the cleaned data.
  - Update the dashboard to view average values for hours or days.
  - Further research in the different usage of students and staff.
  - Create a 'find me a workspace' application for student purposes.
- Future Applications:
  - To implement floor and room SLs, a higher resolution of positioning is required. To achieve this higher resolution the angle of arrival and time of arrival can be used.

2. Introduction

As described in the preface this document provides an overview for the Synthesis Project that is part of the MSc Geomatics for the Built Environment. The project focusses on using WiFi monitoring for collecting patterns of movement to research occupation, to reveal individual movement patterns and flows of people and to discover (related) activities and activity patterns (van der Spek & Verbee, 2016). This report focusses on utilizing WiFi monitoring to determine occupation. The research team exists of five students and is known under the name project in_sight. The research is required to take the following cross cutting topics into account:

- Privacy. How sensitive is the data towards personal information?
- Validity and Accuracy. How reliable and accurate is the data?
- Representativeness. To what ratio are actual users registered across all location?
- System of Access Points (APs). How well equipped is the provided system for monitoring of use?

2.1. Problem Description

The Facility Management and Real Estate department (FMRE) is the manager of the real estate portfolio of the TU Delft. The aim of FMRE is to provide sufficient space at educational facilities for students and staff and is thereby in need of sufficient information regarding the usage of the TU Delft facilities.

Currently, knowledge about the occupation and exploitation of educational facilities is lacking. On behalf of FMRE, Bart Valks reached out to project in_sight to provide them with a solution for this problem. Through analysing data records of the available WiFi infrastructure of the TU Delft project in_sight aims to resolve this problem. Knowledge about the occupation and exploitation aids FMRE in their decision making process.

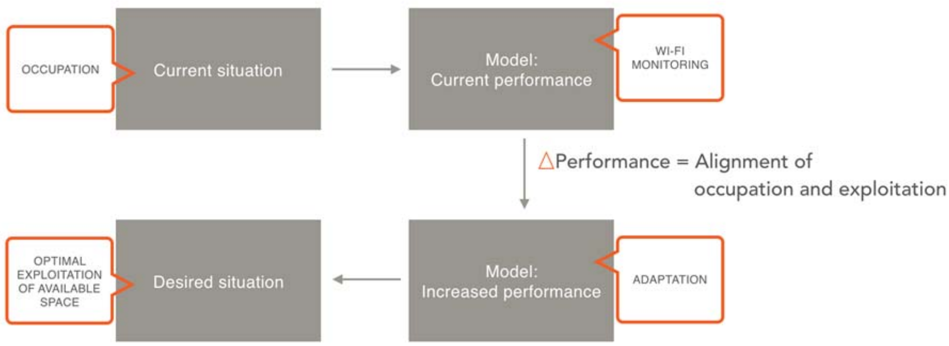


Figure 10: Role of Decision Model, adapted from Binnekamp (2014).

The question to be answered can be defined as a managerial problem, for which a technical solution will be provided (Binnekamp, 2014). This solution is meant to bridge the gap between the current and desired situations, by creating a performance model based on gathered information (Figure 10). The current situation is expressed in occupation, which is mapped through WiFi monitoring. The desired situation is represented by optimal exploitation of available space.

2.2. Research Subjects

In order to derive the occupation and exploitation of educational facilities in the FMRE portfolio, the research to be done has to be clearly defined. This paragraph explains all important subjects to be taken into account, which together build a framework for the research question and scope.

2.2.1. Buildings in Scope

The TU Delft consists of multiple facilities that offer accommodation for both students and staff. There are eight faculties to be distinguished amongst these facilities. The buildings in scope of this project have been listed by FMRE and are as follows:

Building Name	No	Building Name	No	Building Name	No
Science Centre	03	Faculty of Applied Sciences: Biotechnology	05	Faculty of Architecture	08
Faculty of Applied Sciences: Chemical Engineering	12	Aula Conference Centre	20	TU Delft Library	21
Faculty of Applied Physics	22	Faculty of Civil Engineering and Geosciences	23	Faculty of Technology, Policy and Management	31
Faculty of Industrial Design	32	Faculty of Mechanical, Maritime and Materials Engineering	34	Drebbelweg 5	35
Faculty of Electrical Engineering, Mathematics and Computer Sciences	36	Sports Centre	37	Culture Centre	38
Combined Heat and Power Plant	43	Low Speed Windtunnel Laboratory	45	Logistics & Environment	60
Delft Aerospace Structures & Materials Laboratory	61	Faculty of Aerospace Engineering	62	Simona Research Flight Simulator	63
Aerodynamics Laboratory	64	The Fellowship	66		

Table 1: Buildings in Scope. Building Numbers correspond to Figure 11.

Figure 11: TU Delft Campus Map (Technische Universiteit Delft, 2010)

2.2.2. Occupation

The research subject occupation is defined as the number of persons on a Spatial Level (SL) at a specific date and hour. This is an important definition, as the filtering methods, described in chapter 5, are directed to determine the occupation.

2.2.3. Exploitation

The exploitation is defined as an indication for square meters [m^2] per person [n] on a single SL unit. To determine the exploitation rates, the ideal exploitation is calculated by examining the research of (den Heijer, 2011). The ideal exploitation is set to be 14.5 m^2/n (Table 2), which relates to an exploitation percentage of 100%.

	[%]	[m^2/n]
Office space	0.34	20
Educational space	0.18	4.1
Ideal exploitation (calculated)	100	14.5

Table 2: Calculation of the Ideal Exploitation, Based on Data Gathered from den Heijer (2011).

An example of fictive *building A*, which has a UFA of 1.015 m^2, is shown in table 3. The higher the exploitation ratio [m^2/n], the lower the exploitation percentage [%]. This is logical, as a high exploitation ratio relates to much space available per person and thus indicates a low exploitation percentage.

Occupation [n]	Exploitation [m^2/n]	Exploitation [%]
90	11.3	128.3
70	14.5	100
50	20.3	71.4

Table 3: Example Calculation for Exploitation.

2.2.4. Alignment of Occupation and Exploitation

In this research, the alignment refers to the relation between the actual number of persons and the desired number of persons on an SL. It is therefore the alignment between occupation and ideal exploitation. When fictive *building A*, as mentioned in paragraph 2.2.3, has an occupation of n = 70, the desired exploitation of 14.5 m^2/n is reached. Therefore, the alignment is expressed by the exploitation percentage.

2.2.5. SL Definitions

As an SL refers to a specific spatial scale, there are different visualisations, limitations and set-ups linked to each. The setup of an SL is the description of the method in which the data is aggregated, filtered and analysed. Each SL is shortly discussed below.

SL0: Campus

The Campus level includes the occupation and exploitation of the TU Delft campus. It only excludes people travelling across campus as they may not have destinations on TU Delft grounds and if destined to a TU Delft facility, a new record will be made in the *wifilog* as the next building is entered.

SL1: Facility

The Facility level includes the occupation and exploitation of the separate TU Delft buildings. The education facilities in the scope of this project are listed in paragraph 2.1.1. The SL includes people travelling in hallways. It excludes people travelling outside the building, who are considered to be incidentally registered by APs with short session duration. People paying short visits to educational facilities are also removed, since these users do not actively use workspaces or educational spaces.

SL2: Floor

The Floor level includes the occupation and exploitation of a single floor of an educational facility. This includes people travelling in hallways, though hallways are aggregated into their respective floors. Potential problems will be the vertical propagation of AP signals in buildings, and how to deal with sub floor levels. This SL also does not include people travelling outside and people paying short visits to the facility.

SL3: AP

The AP level includes the occupation and exploitation of the AP level of educational facilities. This is the level on which the WiFi monitoring data is gathered and which all former SLs are aggregated from. In order to provide a reliable indication of occupation and exploitation at AP level, the exact location of each AP on the TU Delft campus has to be known. Furthermore, the range of an AP has to be determined, in both a horizontal and a vertical direction. No data will be published on this SL.

SL4: Room

The room level includes the occupation and exploitation of a single room of an educational facility. Ideally, the data from AP level is sub-categorised and filtered to create room divisions. Potential problems are however related to the current unknowns at the AP level. The exact AP location, the exact meaning of signal strength, the influence of building construction and the representativeness of the *wifilog* need to be known in order to derive numbers at SL4. If several APs record the same user, triangulation could be used to locate users in a same room, as well as if a single room contains a single AP with no overlap. The room level can at the very least be examined on a theoretical level and through case studies.

2.2.6. Research Objectives

In order to define the research objectives, the desires of the two main stakeholders of the project are mapped. A complete overview of the stakeholders and available experts can be found in Appendix I. The specific objectives of FMRE and Geomatics can be seen in table 4.

FMRE	• Publish derived data on dashboard. • Nominative division of exploitation. • Preferably campus wide implementation for SL0, and Industrial Design (new AP system) versus another building with an old AP system for further SLs. • Proven reliability percentage for outcomes.
Geomatics	• Rational division of exploitation. • Develop conceptual dashboard for each SL. • Proven reliability for outcomes.

Table 4: Mapping the Objectives of the Main Stakeholders.

The objectives for the research are as follows:

- Determine the occupation of facilities on the TU Delft campus accurately by analysing WiFi connections of Eduroam users.
- Calculate the exploitation of facilities on the TU Delft campus on basis of the gathered occupation.
- Create an online dashboard, which is accessible for the main stakeholders, that visualizes the research findings in an interactive manner.

2.3. Research Question

The problem description introduces the knowledge gap for FMRE. The company has access to the *wifilog* of Eduroam users available in a database. As the purpose of the Eduroam system is to provide a WiFi connection to its users, the system could aid in determining the occupation and exploitation of TU Delft facilities. The project is initiated by the idea that communication technologies can be used to collect information. To answer this, the research question is defined as follows:

To what extent can the alignment of occupation and exploitation of educational facilities on different scales be indicated through WiFi monitoring?

The research is specifically focused on the presence of people in sense of occupation and exploitation of available space and to estimate the number of devices at a location at a certain time, representing presence of people at that place at that time or for certain duration.

2.3.1. Hypotheses

In order to answer the proposed research question, an overall hypothesis is defined and divided into sub hypotheses. As the final output of a hypothesis may be either positive or negative, its direct cause is defined through the definition of a set of sub hypotheses. The research tests the following hypothesis:

H1: The alignment of occupation and exploitation of educational facilities **can** be indicated through WiFi monitoring on all SLs with adequate reliability.

H0: The alignment of occupation and exploitation of educational facilities **cannot** be indicated through WiFi monitoring on all SLs with adequate reliability.

The three sub hypotheses that contribute to answering the overall hypothesis are the following:

	H1	H0
1	The occupation of facilities can be determined with **80% or more** reliability.	The occupation of facilities can be determined with **less than 80%** reliability.
2	The occupation and exploitation of use types 'educational' and 'staff' **can** be determined.	The occupation and exploitation of use types 'educational' and 'staff' **cannot** be determined.
3	The available system of APs **is** suitable for indicating occupation of the TU Delft campus through WiFi monitoring.	The available system of APs **is not** suitable for indicating occupation of the TU Delft campus through WiFi monitoring.

Table 5: Sub Hypotheses and their Potential Outcomes.

2.4. Research Scope

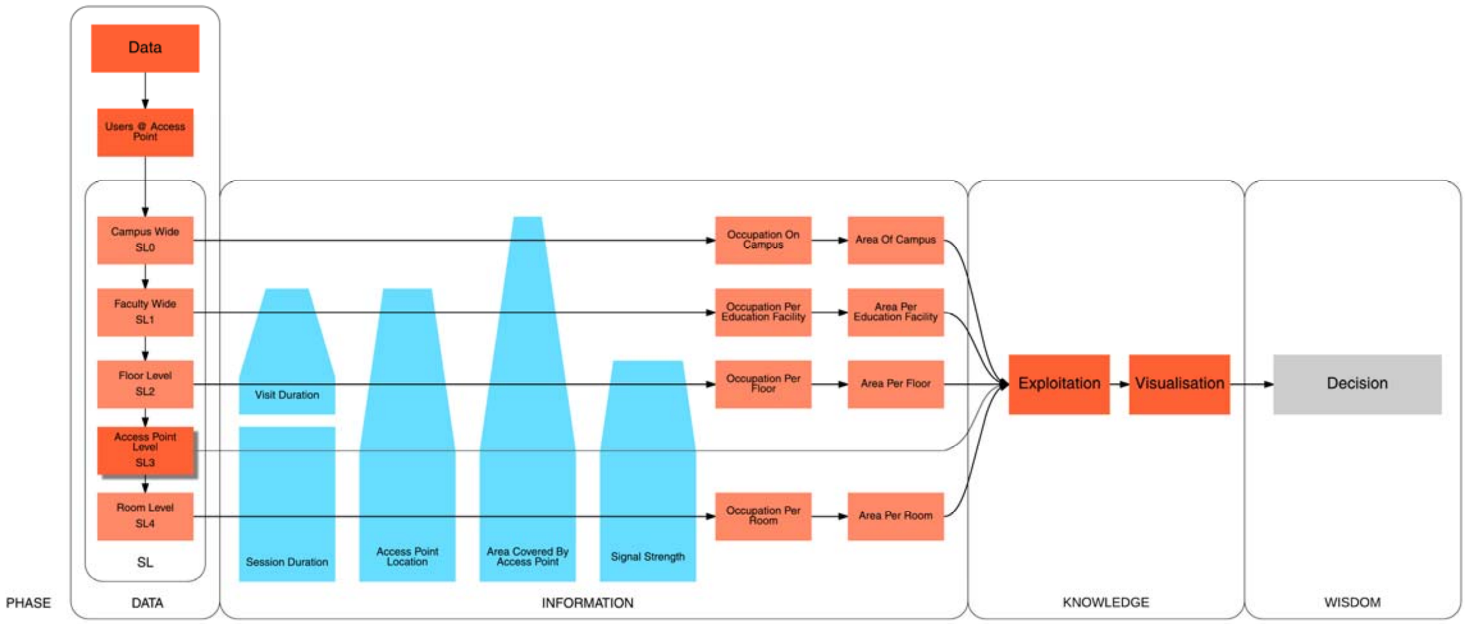


Figure 12: The Project Scope with the Information Flows Visualized in Blue (Appendix II).

Based on the research question and the hypothesis definition, the scope of the research is defined (Figure 12). Generally, the process is divided into four stages: data, information, knowledge and wisdom. Each new stage represents a new level of understanding of the content of the data, by connecting and analysing all data and information available (Ackoff, 1989). This to solve the general problem of handling raw data, as the *wifilogs*, in which context and relations to other data are not present (G. Bellinger, D. Castro, & Mills, 2004)

2.4.1. Data

The data stage contains all available sources as input for the project. The available data sources are listed in table 6. Before connection, the data is divided into SLs, as described in paragraph 2.2.5.

Data Content	Data Type	Data Source
wifilog (raw WiFi monitoring data)	PostgreSQL table	ICT TU Delft
Buildings in scope	CSV	FMRE (Valks)
UFA per room, per building	CSV	FMRE (Valks)
Campus Map (outline, height)	Shapefile	BAG
Map per building floor with use type	PDF	FMRE (Valks)
Validation Data	CSV	FMRE (Van Duinen, Mantel)

Table 6: Available Data Sources.

2.4.2. Information

Information can be directly or indirectly derived from the data, and can be used to aggregate sets of data into meaningful groups. Since the type of useful information may differ for each SL, the influence of the defined information flows on the research scope are defined per SL. Based on the influence of each information flow, the occupation can be calculated from the filtered and aggregated data. These data handling strategies are described in chapter 5.

Session Duration

Session Duration is defined as the time in which a user is recorded by a single AP. An aggregate for this value is the Visit Duration, as the time in which a user is recorded in a certain area, dependent on the SL at stake. The influence of sessions and visits is defined in table 7.

SL0	Not applicable. People travelling across campus will be counted, as they most likely have a destination on campus. Guest travellers do not influence campus occupation, as they do not have an Eduroam account.
SL1	Logged sessions for users within the same building and of at least fifteen minutes are aggregated as a visit.
SL2	Logged sessions for users on the same floor are aggregated as a visit. If the visit duration is less than ten minutes, the user visit is classified as travel time and filtered out.
SL3 **SL4**	Each user is recorded in sessions per AP. Sessions shorter than five minutes are classified as travel time and filtered out.

Table 7: The Influence of Session Duration for each SL.

AP Location

The AP location indicates the location of the user, though the precision and correctness of these locations can only be checked manually. Maps of the AP locations would aid this information flow, but these are not available. The influence of AP location is defined in table 8.

SL0	The exact location of APs does not influence campus level.
SL1	The APs are grouped by facility based on the attribute *apname*.
SL2	APs are grouped per facility, by floor level. Every facility requires a unique handling of AP grouping per floor, as some facilities clearly record this information in *apname* and / or *maploc.*
SL3	As records are registered by AP adjacency, the exact location of each AP is high influence.
SL4	The vicinity of a record to an AP should determine in which room the recorded user is. However, this translation requires additional information about the reach and direction of an AP signal.

Table 8: The Influence of AP Location for each SL.

Area Covered by AP

Part of the research is aimed at the reach of each AP, in both the horizontal and vertical direction. Furthermore, the actual coverage may be influenced by external influences, like wall placement and their thickness. The AP records contain information about the reach of each AP, since they are in flux. In the case an AP fails, the ones around it are programmed to increase in signal strength and therefore 'catch' the area originally covered by the broken AP (Cisco, 2015).This particular setting increases the complexness for the project. The influence of the AP coverage is defined in table 9.

SL0	Available APs are expected to cover all relevant facilities, but outside area is covered as well.
SL1	Available APs are expected to fully cover each facility.
SL2	APs might reach across floors, so signal strength may be useful to filter on.
SL3	A map depicting reach per AP may show whether the AP system covers full facilities, and can gives a concise representation of occupation.
SL4	Area coverage of APs is expected to go beyond room level, and therefore requires additional information.

Table 9: The Influence of Area Covered by AP for each SL.

Signal Strength

One of the attributes to each record is the signal strength of the connection. A shortcoming in the system is the inconsistency of RSSI recording, where some values are exceptionally high (-10 dBm) and others are in a seemingly false value (-128 dBm). Despite these inconsistencies, the signal strength may be used to enrich information anyways. The influence is defined in table 10.

SL0 **SL1**	Not applicable. Derived values for these SLs are aggregated values for relatively discernible sets of APs per facility.
SL2	Low signal strength might mean two things: the recorded user is on a different floor, or the user has moved, but is still connected to a different AP due to roaming settings.
SL3	Signal strengths indicate adjacency to AP. However, the stored signal strength values are an average value for the five-minute record.
SL4	As the recorded signal strength is only an average value and contains no directional information whatsoever, the signal strength requires additional information.

Table 10: The influence of Signal Strength for each SL.

2.4.3. Knowledge

The finalization of the processing of all numerical data is made in the calculations for exploitation. These values are then visualized on the dashboard, as described in paragraph 7.2.1. The dashboard utilizes the specific spatial partitions of the data, so that all produced outcomes can be visualized in a 3D model.

2.4.4. Wisdom

As the wisdom stage entails decision-making, this process is left to the project client. In order to answer the question imposed by FMRE, the project outputs a dashboard as Spatial Decision Support System (SDSS). FMRE can use the occupation and exploitation values published on the dashboard as information flow to be connected with additional data sources available to them.

2.5. Reading Guide

The following chapter describes the methodology for the conducted research. In this chapter the stakeholders, requirements, organization and planning will be discussed. Chapter 4 discusses the data capture process. The WiFi technology and theory will be described. Chapter 5 focuses on data analysis. In this chapter, the reason for cleaning and the different cleaning steps will be described. Chapter 6 describes the storage of the processed data. Chapter 7 describes the data visualization. Chapter 8 focuses on data quality control throughout project. Chapter 9 handles the cross cutting topics defining the bounds of the project. Privacy, validity and accuracy, representativeness and the system of APs will be discussed here. The report is concluded with an evaluation, conclusion and recommendations in chapters 10, 11 and 12.

3. Methodology

This chapter captures the goals, constraints, requirements and risks that are to be taken into account in order to define the research methodology. The rich picture, the requirement analysis, the risk register and an overview of the project organisation are therefore discussed and used as input to define the used methodology.

3.1. Rich Picture

The first step in exploring the requirements and stakeholders in the research project is to create a rich picture (Figure 13). The visualisation provides an overview of the subjects to be considered in the research. Four phases can be distinguished: data capture, storage, analysis and communication.

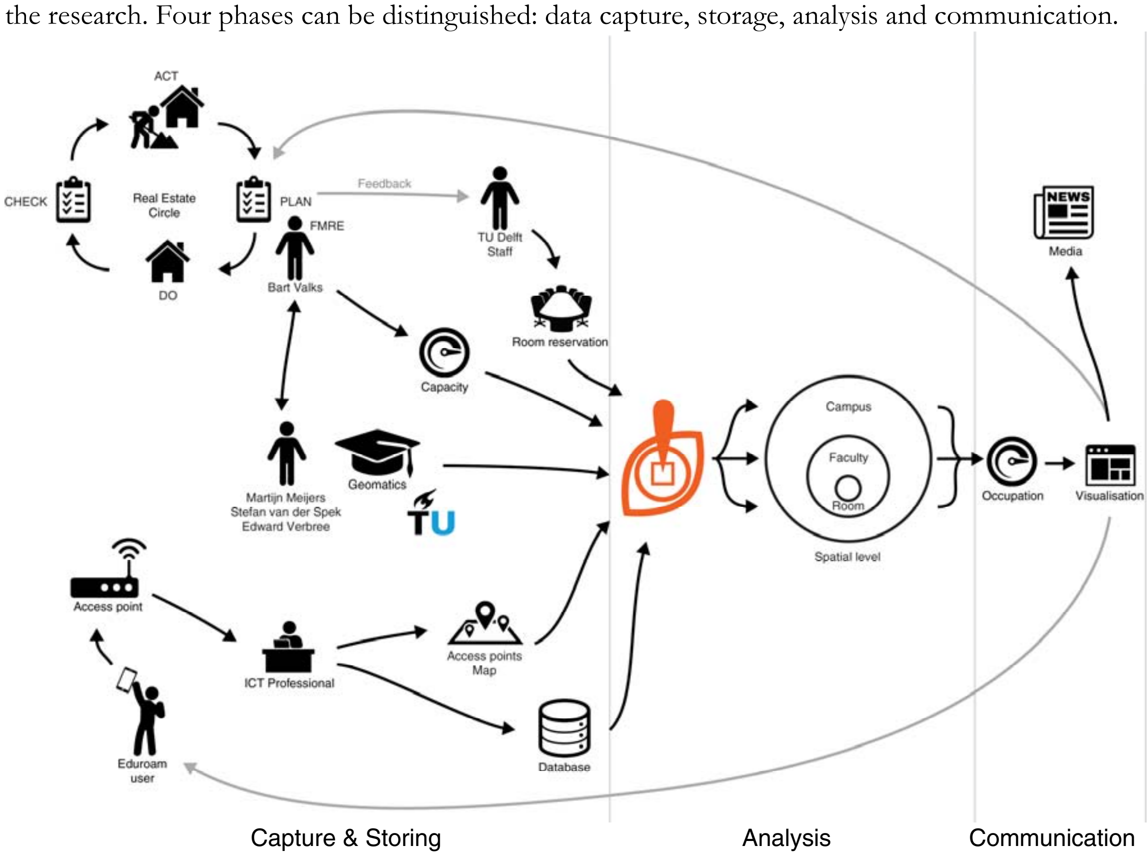


Figure 13: Rich Picture, Showing Stakeholders and Resources involved, divided into subsequent Methodological Phases.

3.2. Requirement Analysis

The entire process is highly influenced by the requirements from the stakeholders involved. Throughout the process, both conclusions and products are tested to the set requirements in order to stay on track. Though these requirements have been defined at the beginning of the project, they may change over time. A change in requirements may be classified as a risk and its effects are described paragraph 3.3. This paragraph describes the project requirements as well as constraints encountered during the process.

3.2.1. Client Requirements

The main client, FMRE, aims to improve their understanding of the use of the campus, as to develop new strategies for optimal use of their portfolio. FRME expects an estimation of the efficiency of the 26 buildings within the scope, in order to be used for predictive occupation. The client requirements can thus be summarized as follows (Valks, 2016a, 2016c):

- A focus should be on the amount of users, students and staff, in TU Delft facilities, with special interest in the facility and floor levels.
- The knowledge gained from the project should be usable to develop portfolio interventions.
- FMRE prefers an online dashboard as output.

3.2.2. Course requirements

As the project is part of the MSc Geomatics programme, specific course requirements are to be distinguished. The overall research direction and focus on cross cutting topics stem from the course requirements, as well as specific topics described within the reports. The course requirements are elaborated in the project guide (van der Spek & Verbee, 2016).

- The research project encompasses the different disciplines as taught in the MSc Geomatics for the Built Environment at TU Delft.
- The research team should describe the entire geo-information process from project definition, measurement, data processing and analysis, to presentation, delivery and application.
- The research team should actively collaborate internally, but also seek to contact staff and professionals.
- The research project encompasses an innovative project that extends 'state-of-the-art' with new fields to explore.

3.2.3. Functional requirements

The functional requirements describe the features, functioning and usage of a product from the perspective of the product and its user. The output of this research is project should be a product that contains the following:

- Online dashboard, which should provide insight into the occupation and exploitation of the TU Delft campus.
- An SDSS, which should provide knowledge to aid decision making for asset management on different SLs.
- Automatic updates of views and tables when new data is added to the *wifilog*.
- Historical data, which provide insight into occupation on past dates.

3.2.4. Non-Functional Requirements

Non-functional requirements describe the qualitative factors that affect the functionality effectiveness. The output of the project is determined to be a product that contains the following:

- The data is available with a temporal resolution, which has a preferred range of fifteen minutes.
- Availability of the online dashboard, which spans beyond the duration of the project.
- Accessibility, which should be without the use of a password or downloading of additional software.
- A quick response time of the interface, which makes the dashboard user-friendly.

3.2.5. Constraints

Where the requirements guide the project in a general direction, the constraints define the outer bounds of the research fields. These constraints need to be defined in order to assess risks to the project, as well as foreseen limitations.

Killer Requirements

The requirements that could endanger or obstruct the feasibility of the project are identified as follows:

- Limited number of APs. In most cases, multiple rooms share an AP, which could be a limiting factor to the accuracy of occupation calculations and visualisations on certain SLs.
- Limited measurement information. The WiFi system only records Eduroam users, and skips campus visitors or those who choose to opt out. Signal strength is measured as an average value for the duration of a single session, and gives no actual information on where the recorded user could be.
- Data sources. The project heavily relies on external data sources. If the final product should work after completing the project, these sources need to be updated automatically.
- Privacy. The misuse of personal data recorded in the WiFi tracking data should be prevented. The usernames and mac addresses of all users and their devices are hashed. Protests of students or staff could obstruct the feasibility of the project.

Further Constraints

In the process so far, further constraints have arisen, limiting the possibilities of the project.

- AP locations. The TU Delft ICT department has promised to provide maps with the locations of the APs within each facility. However, these maps have not been provided yet.
- Roaming function. The APs are set up by the manufacturer to always provide the user with the best possible internet connection (Cisco, 2016). Though a desirable setting for the user, this means often times a user is not connected to the closest AP.

3.2.6. MoSCoW

The MoSCoW rules are a means for project analysing and prioritizing (IIBA, 2016). Within this method, all Must requirements have to be fulfilled in order for the project to succeed. The definition of Won't requirements encapsulate the project by defining subjects that may relate to the project by defining subjects that may directly relate to the project, but will not be researched.

Must	Should
• Contain sufficiently accurate information on the device connections to the APs to facilitate positioning on multiple SLs. • Realise an online dashboard for easy access by FMRE to the processed data. • Have at least one SL covering the campus. • Have at least two SL for analysing, filtering and visualising the data. • Give FMRE insight into problem areas on the campus by using filtering tools. • Have the implementation for historical data. • Contain research on the capabilities of the APs with the focus on positioning persons and room and floor boundaries.	• Have a system that can be accessed by FMRE after the completion of the Synthesis project. • Contain research on the range of the APs and the signal strength propagation. • Have analysis of chosen research and visualization tools.
Could	**Won't**
• Have partial implementation of SL3-4. • Have an application focussed on the student as a stakeholder. • Have campus wide coverage of SL0-2. • Have a 3D visualization of select SLs.	• Have campus wide implementation for SL3 and SL4. • Monitor the attendance of employees based on the exploitation of office space. • Provide a solution to the exploitation challenge for FMRE. • Have the implementation for real-time data.

Table 11: MoSCoW Rules.

3.3. Risk Register

A risk can be defined as the probability of an adverse outcome and the severity of the consequences if the outcome does occur. In this section, the aim is to elaborate on the important risks in the research project. Focus is on the three steps in handling risks: Identify, analyse and respond. A first step is to identify possible risks. Subsequently, the probability and impact of the risk has to be analysed. The final step is to define a proper risk response to handle the risk. Examples are described by the ATOM method, which describe to Accept, Reduce, Avoid or Transfer risks (Hillson & Simon, 2007).

The objective of risk identification is the early and continuous identification of events that have negative impacts on the project, if they appear. These impacts may be on performance, outcomes or goals defined. In this research project, the risk identification is focused on the following aspects:

- Technical assessment, as system validity and reliability through representativeness of users, APs and accuracy of data.
- Organizational backgrounds.
- Privacy, as the sensitivity towards personal data.
- External grounds, mainly by the involved stakeholders.

Three technical risks are elaborated on in table 12. The full Risk Register can be found in Appendix III.

		Risk 1	Risk 2	Risk 3
Risk description	Cause	Automatic / manual validation of provided data not possible within time span of project	The map of AP locations cannot be provided by the ICT specialist	Right to object of students/staff to opt out to the Eduroam system
	Risk event	No accurate conclusive information about the quality of the data output can be provided	The AP locations are not available on a map	Manually eliminate data entries of students/staff
	Consequences	Reliability of data output cannot be guaranteed to client	Mandatory decrease in scope due to manual validation of AP locations	Decrease of data quality
	Project promise	Quality	Time, Scope	Quality, Time
Pre response assessment	Probability	MODERATE	HIGH	LOW
	Impact	VERY HIGH	VERY HIGH	LOW
Risk response	Description	Manual validation at an early stage, foresee client with issues regarding this risk and involve client in validation process	Discuss with mentor and client the issues regarding this risk towards the scope and validation of the project	The few Eduroam users that are aware of this project and that do object can be taken out of the database
	ATOM	Reduce	Transfer	Accept
Post response assessment	Probability	LOW	LOW	MODERATE
	Impact	VERY HIGH	VERY HIGH	LOW

Table 12: Overview of three Technical Risks in the research project.

3.4. Organisation

The synthesis project is a real-world assignment with multiple stakeholders and students for a period of 8 weeks. To have a clear understanding of the responsibilities for each individual student an organisational breakdown structure (OBS) is created which will be discussed in this chapter. After the description of the OBS a global overview of the different tasks will be given. The paragraph will end with the explanation of the weekly process.

3.4.1. Organisational Breakdown Structure (OBS)

The research project can be split into two areas of interest (Figure 14). First, FMRE is the commissioning party and seen as the main client of the project. Second, the Geomatics department as research group is demanding a research project that aligns with the Synthesis Project course description. An extensive elaboration of the OBS can be found in Appendix IV.

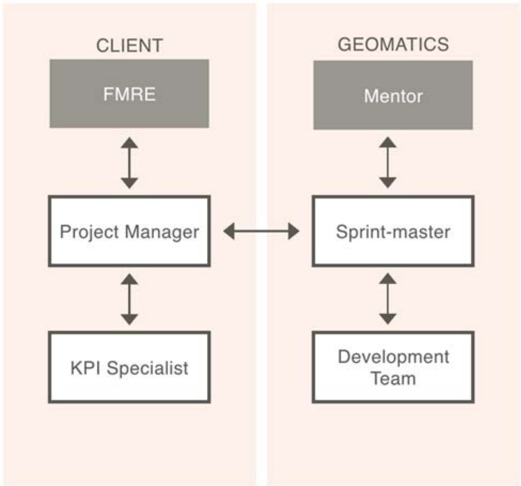


Figure 14: Organizational Relations within the Research Project.

3.4.2. Work Breakdown Structure (WBS)

To have a clear overview of the different tasks the previous chapters are summarized in a WBS. This schema gives a clear insight into the tasks that need to be performed (Figure 15). The translation into practical tasks can be found in the sprint execution documents (Appendix V). The WBS itself can be found in Appendix VI.

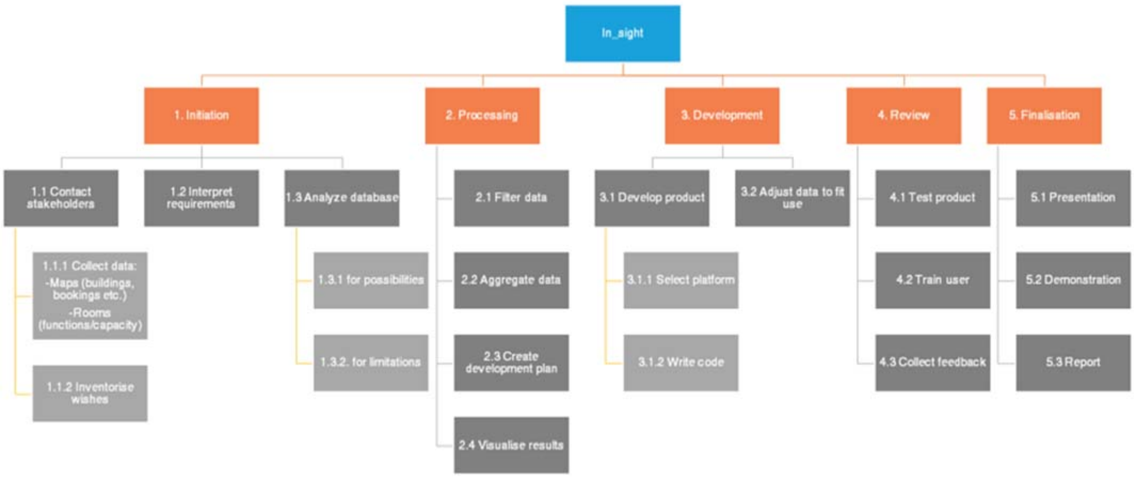


Figure 15: Work Breakdown Structure.

3.4.3. Sprints

In order to make a design process both efficient and productive, it can be performed in design sprints (Google, 2015). A design sprint can be divided into six stages (Figure 16). These stages allow the process to be diverged three times, always followed by a converging stage. This improves the quality of the research and design process, by challenging the project team to explore all possible options and clearly define the research direction constantly throughout execution. An elaboration on the stages and the practical sprint execution can be found in Appendix V.

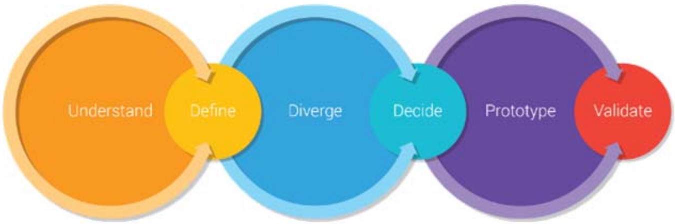


Figure 16: Stages of a Design Sprint (Google, 2015).

In order to keep the project process streamlined, each phase is divided into design sprints. As the first phase is only a week, this week is used for defining the project and final products. Both the second and third phase is used to test hypotheses to the SLs that will be dealt with during the sprints (Table 13). The fourth phase is used to finalise the results for the final report and Symposium.

	SL0 (Campus)	**SL1 (Faculty)**	**SL2 (Floor)**	**SL3 (AP)**	**SL4 (Room)**	**Evaluation hypothesis**
Hypothesis	Sprint 1	Sprint 1	Sprint 2/4	Sprint 2	Sprint 3/5	
Sub-H 1	Sprint 1	Sprint 1	Sprint 2/4	Sprint 2	Sprint 3/5	
Sub-H 2	Sprint 1	Sprint 1	Sprint 2/4	Sprint 2	Sprint 3/5	
Sub-H 3	Sprint 1	Sprint 1	Sprint 2/4	Sprint 2	Sprint 3/5	
Evaluation of SL						

Table 13: Division of SLs and Hypothesis Assessment for the Design Sprints of Phases 2 and 3.

3.5. Research Methodology

The previous paragraphs presented an overview of the relations in the Rich picture, an extensive requirement analysis, the identified risks in a risk register and the organisational relations. These are used as input to define the methodology. An overview of the methodology is shown in figure 17. The methodology steps are based on Lemmens (2011), and describe the six stages that are important when handling geo data. First, the data capture provides theory about WiFi and the data storage of the raw data in the available WiFi monitoring system. The data analysis presents insight in the filtering methods that are used to filter, group and aggregate data to the desired output in occupation and exploitation. The next step is to describe the data storage of the filtered data. The data structure provides an explanation of the resulting tables and views in the database. The communication and visualisation includes a description of the interface and the abilities of the dashboard. As the validation of data quality is important in all previous steps, multiple case studies are performed. Last, the management step refers to the decision-making as is to be performed by FMRE.

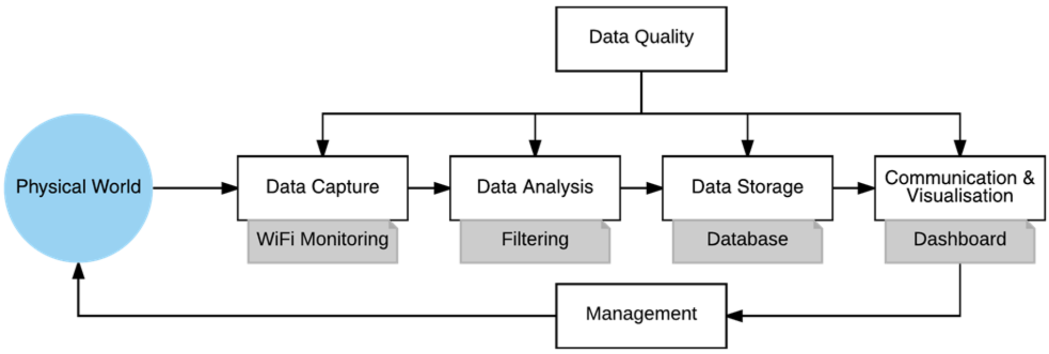


Figure 17: The research methodology, based on the geo-processing steps of Lemmens (2011).

4. Data Capture

The sensors used in this project are APs, which are used to act as sensors that record the connected users. However, their primary purpose is to provide an internet connection to all requesting devices. To understand more about the ability of APs to function as a sensor and to be able to capture data, concepts about WiFi signals, system concepts and data storage of raw data are discussed in this chapter.

4.1. WiFi systems

WiFi signals are Electromagnetic (EM) signals and are the result of electricity flowing through metal antennas. These waves travel into space, away from the emitting antenna. Their range depends on the energy of the emitted signal and the signal frequency. The general range of WiFi signals is greater than Bluetooth signals and between 50 and 100 m (Mautz, 2012). An EM signal can be absorbed by physical objects, like moisture particles and walls. A signal with a higher frequency gets absorbed better by physical objects. Typical frequencies for WiFi signals are signal around 2.4 GHz and 5 GHz. Figure 18 shows an overview of the EM spectrum and the part that of the spectrum that WiFi uses in black.

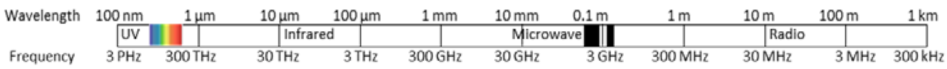


Figure 18: The EM Spectrum with the WiFi Frequency Range (black). Reproduced from Mautz (2012).

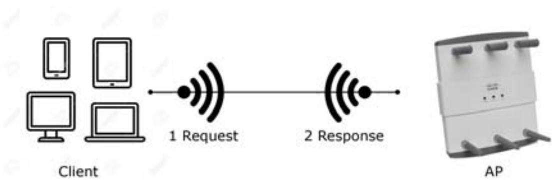


Figure 19: The local AP infrastructure, the active sensors, and the signal transmission between the AP and the client

The 2.4 and 5 GHz frequencies indicate a span of frequencies, a range in which the frequency can vary. These ranges are called the frequency bands. Within the used bands, the subdivisions are called channels. Channels should be used to avoid overlap or interference. Besides physical objects, the signal can also be reduced by other signals. This is especially the case in very crowded areas, which are prone to a lot of noise. The WiFi signal at 2.4 GHz mainly faces this problem, as the same frequency is used by different types of electronic equipment, such as wireless (dect) telephones and remote controls.

Network connections use a two-way communication (Figure 19). Therefore, every receiver can also transmit and vice versa, making every AP a transceiver (Commotion Wireless). Since the most common WiFi signal is in the 2.4 GHz band, the chance for congestion is the biggest when using this type of connection. The 5 GHz signal does not face this problem, as it is supported by fewer devices. However, the range of this signal is shorter and it cannot travel well through floors and walls. WiFi Monitoring. When too many clients connect to a single AP, the Cisco system will bounce clients to other APs, which also causes problems in reliability of the localisation. The available AP system can be used for active monitoring. The APs or beacons actively emit signals. According to Mautz (2012) an active system requires locally deployed beacons to gather data, the system depends on a local infrastructure. On the other hand, a passive system would not require any external infrastructure components and it is able to operate autonomously.

4.2. AP System

The TU Delft is equipped with an AP system delivered by Cisco. The exact workings of the specific system and routers set up on campus determine the quality of gathered data for the proposed research.

4.2.1. Eduroam introduction

Eduroam is shorthand for education roaming. The purpose of Eduroam is to offer students and academic staff internet access at all locations that have Eduroam implemented. To implement an Eduroam network there are some general guidelines and agreements that make sure the network functions consistently across different educational facilities. The most important agreement is the wireless network name (the so-called Service Set Identifier, SSID), that is always Eduroam. The way users authenticate is also standardized. The authentication is always handled by the Identity Provider (IP) and uses IP specific authentication methods. This way the system knows who tries to connect to the network. Once authenticated, the user can be authorized (granted access) by the Service Provider (SP). This role is fulfilled by a WiFi AP itself. Between organisations, there is always a link with a special server, called a RADIUS server that enables authentication of users.

4.2.2. The Eduroam AP System

The TU Delft Eduroam network makes use of professional grade APs of the Cisco brand. The AP type commonly used on this network are APs of the Aironet 1250 product series of Cisco, but these are relatively old and End-of-Sale. Some facilities use newer types, but the same principles apply. Figure 20a displays an AP that is installed on the ceiling of the ground floor of the Faculty of Architecture. This AP is certified by the WiFi alliance (Cisco, 2014b 2008), which ensures that most devices are compatible with the system. These APs can use two different types of antennas. It is a modular system, so the system operator can fit different antennas or radio modules (2.4 GHz or 5 GHz) to the slots on either side. Only the middle part, the band with the LED indicators, will always be the same in every configuration.

Figure 20: Picture of a TU Delft AP Installed at Architecture (a) and one at EWI (b). C: A radio Module that can be attached to the Core Body of the AP, to modify its Behaviour and Characteristics. Reproduced from Cisco (2013).

The circular external antennas on the left (Figure 20a) are broadcasting on the 2.4 GHz wavelength. The flattened antennas on the right are broadcasting on the 5 GHz wavelength. The performance specifications of this AP are according to the draft standard 802.11n, version 2.0. This is an extension on the original 802.11 WiFi standard. Devices that follow this standard can use both the 2.4 GHz and 5 GHz bands. Figure 18C shows a radio module with antenna connectors that can be fitted to the core chassis of the AP. This allows the technician to adjust the AP's characteristics to the environment and/or target users.

Another type that is used at TU Delft facilities is the Aironet 2600 AP as shown in figure 20b. This type does not use external antennas, but also has dual-band functionality. The antennas are integrated, but the signal radiation pattern is comparable.

4.2.3. Technical Background

To transmit data via the Radio Frequency (RF) signal the signal needs to be modulated. There are two main types of modulation for WiFi signals: Direct-Sequence Spread Spectrum (DSSS) and Frequency-Hopping Spread Spectrum (FHSS). It goes beyond the scope of this research to check the differences in reliability of WiFi monitoring between the two types of modulation. Moreover, the Cisco APs only use DSSS modulation as it is cheaper, faster and has better range. Generally the range of WiFi APs is between 50 or 100 m, but the actual range depends on the propagation environment in which the AP is placed (Mautz, 2012).

Antenna properties

For this research, it is important to know how antennas are placed and what their reach is in all directions as this can say something about the reliability of the WiFi monitoring results. Three fundamental properties are influencing the reliability: the polarization, gain and the direction. The direction determines the spread of the signal inside a room, while the gain is a measure for the increase in power of the EM signal by the antenna. The gain is a value in dB relative to a theoretical antenna with even spread in all directions. A higher gain means the signal is more focussed to a specific direction and thus has a smaller coverage. A visible WiFi router antenna normally has a donut-like radiation pattern as show in figure 21. These antennas are called dipole antennas.

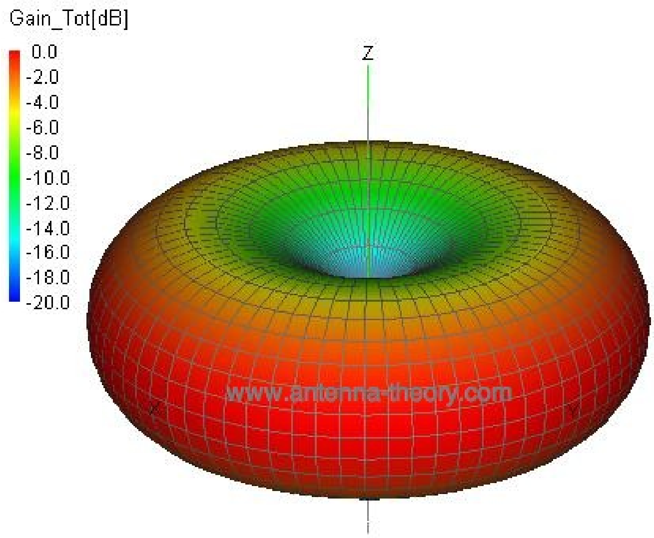


Figure 21: Radiation pattern of a dipole antenna, figure reproduced from antenna-theory.com (2015).

The radiation pattern clearly shows that the signal direction is greater in the horizontal plane than in the vertical plane. As such, an AP will spread its signal mostly inside a room if the antenna is placed vertically. Different types of antennas can also be used to direct the signal in a particular direction. An external dipole antenna can be aimed at a different angle so that the signal can be directed in a specific manner. Figure 13 shows the type of antennas used within the TU Delft facilities, dipole antennas. Because the antennas have an elevation of 65 degrees, some signal will penetrate through the floor. People connected above the floor will be localised. The integrated antennas have a similar effect, albeit fewer signals go upward (see figure 22).

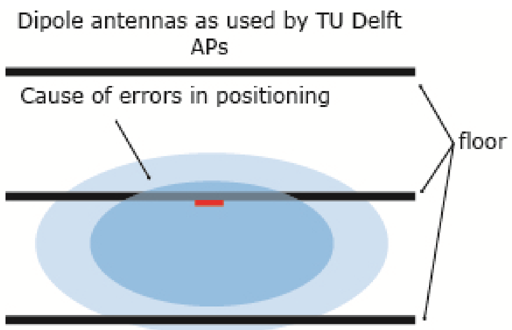


Figure 22: Dipole antennas and signal propagation

The directions, horizontal and vertical, are called azimuthal and elevation planes. The signal direction can be visualised in 2D as depicted in figure 23.

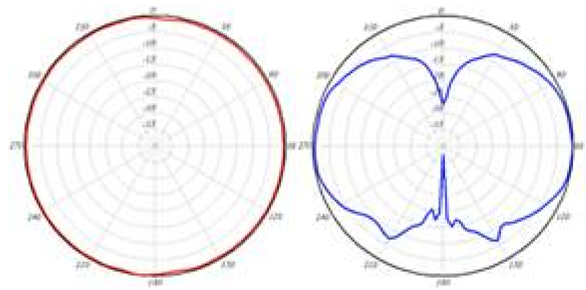


Figure 23: Dipole antenna signal spread, figure reproduced from Cisco (2014d 2014)

Dipole antennas are chosen because they can best serve the user (with a wireless transmitter). This is due the fact that the location of the user is not known beforehand and thus homogeneous coverage for the whole room is the best solution, but the signal should be focussed on covering one floor. The APs at the TU Delft facilities are always mounted to the ceilings with their external (if present) pointing downwards so the radiation pattern should look like that of a horizontal donut. The actual shape of the radiation pattern depends on the specific situation. Some AP types used in the TU Delft network use integrated omnidirectional antennas. The radiation pattern of these APs differs slightly from that of the external dipole antennas as shown in figure 24. The major difference is that the signal radiation pattern also looks like a donut in 3D, but the centre of the donut is below the AP itself, so less signal past the bottom of the AP.

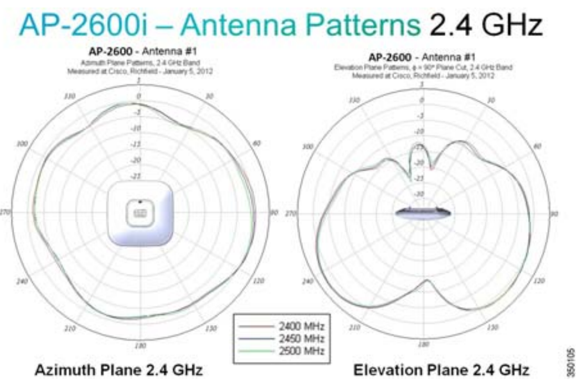


Figure 24: Radiation pattern of the APs with integrated antennas, figure reproduced from Cisco (2014c)

Coverage and connections

To be able to monitor the occupation in every room the whole area of a building should be covered by WiFi signals, to achieve this multiple APs must be placed. The APs can be interlinked by network cables. Generally, a signal can penetrate one or two walls before the quality is degraded too much. Signal penetration also depends on the thickness and density of the materials. Floors often contain steel reinforcement and therefore the penetration is less.

apname	time	nr	date
A-62-0-018	13:13:00	171	22-4-2016
A-62-0-018	12:47:00	167	22-4-2016
A-62-0-018	13:08:00	166	22-4-2016
A-62-0-018	13:03:00	158	22-4-2016
A-62-0-018	13:18:00	156	22-4-2016

Table 14: Maximum number of users connected to a single AP

The number of clients that can be connected to a single AP is limited to 200 simultaneous users. Our tests have indicated that this number is never met. Even on times where there are special events (for example BK Beats), the maximum number of users is not reached. As table 14 shows, the AP with name A-62-0-018 has 171 connected users at 13:13 on 22-04-2016.

4.2.4. Data capture problems: other connections

This section highlights some of the problems with the used infrastructure for calculating the occupation and exploitation.

TU Delft Visitors

People within the building cannot always be tracked through the current AP system. Visitors do not connect with the Eduroam network, as they do not have an Eduroam account with which they can authenticate with the system. Instead, visitors connect to a special visitor network, TUvisitor. This connection also uses the AP system, but the connections are not recorded in the *wifilog* database. A recommendation for further research and development is to include also these connections in the *wifilog*, with the addition of a column specifying the used network name.

Local Area Network

Another group of users that is not currently recorded in the *wifilog* table are the users connected to the network by an Ethernet cable. Testing shows that a user can be tracked when he has WiFi enabled. When the user switches to a LAN connection, the mac address will change to that of the LAN network adaptor. However, the username is not recorded anymore in the *wifilog* (appendix X) As shown in Figure 25 these users do not connect through the APs, but indirectly to network switches via wall sockets. WiFi monitoring is not an option for these users, however the approximate location of a user can be known if the port on the switch is saved together with the connection data and the location of the wall socket leading to the port on the network switch.

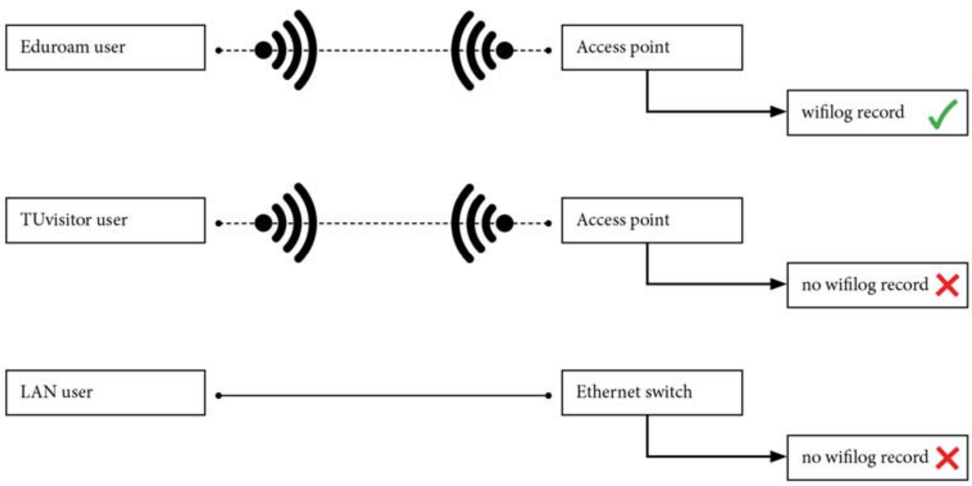


Figure 25: Different connections: the users of the TUvisitor and the LAN network are not recorded in *wifilog* and thus cannot be used monitor occupation.

4.3. Data Storing

The provided data is based on the *wifilog*, gathered by each separate AP. The TU Delft ICT department stores these records as raw data on a database, with remote access provided to the team.

4.3.1. Data Structure

The raw data used is referred to as the data, or the database table *wifilog*. A single entry within this database is referred to as a record and contains the following information:

- *Username*. A hashed code as a unique identifier for each Eduroam user.
- *Mac*. A hashed code as a unique identifier for each connected device.
- *Asstime*. The timestamp a specific record is associated with. It is the starting time of a session and contains of a date of format yyyy-mm-dd and a time of format hh:mm:ss.
- *Apname*. A TU Delft specific name for the AP in question of the following format: A-X-Y-ZZZ, where X stands for building number, Y stands for floor or wing in select faculties and Z stands for a specific AP number.
- *Maploc*. A description of the AP location with building number, building name and floor, which may be as specific as "ground floor+" or as unspecific as "2nd, 3rd and 4th floor". Many of these descriptions contain spelling mistakes and discrepancies.
- *Sesdur*. The duration of the recorded session, on a time interval of around five minutes.
- *SNR*. Signal to noise ratio as a measure of signal quality emitted from the AP.
- *RSSI*. Signal strength that a device receives from the AP in dBm.
- *Importfile*. A specification to which file added the data on which data.

5. Data Analysis

Data analysis is an essential step when going from data to information (Ackoff, 1989). An analysis does not consist of a single large step, but several small steps that ultimately lead to a result, which can then be visualized to provide information to the stakeholders. In this chapter, the focus is on the small steps, which eventually lead to results that can be used for visualization. First, the value of filtering data is discussed. Next, the definition of a visit and its use in filtering is described. The queries used to create the different tables in the database are discussed in paragraph 5.3. Paragraph 5.4 discusses the effect of the queries is discussed.

5.1. The need for filtering

The data from the *wifilog* table is filtered because each dataset has different characteristics and is therefore in most cases not directly usable for certain applications. The right fitness for use can be achieved by cleaning and filtering the delivered dataset (Lemmens, 2011, p. 32).

There are three groups of records that will be filtered for all SLs: static devices, users living on campus and traffic. Static devices are unmoving objects that are always turned on during daytime. They are characterized by having a large amount of *macs* connected to a single user. By focusing on the high number of *macs* as well as a daily presence, user account with only macs can be identified (Figure 26). Another possible source of noise in the data is students who can access the Eduroam WiFi network from their home. Students connecting from their homes can lead to a higher occupancy rate. These users will make use of the WiFi network after closing time for multiple days (figure 27) with long session durations on a single AP. The traffic aspect differs from the other aspects; the interpretation of traffic changes throughout the different SLs, as will be discussed in paragraph 5.2.

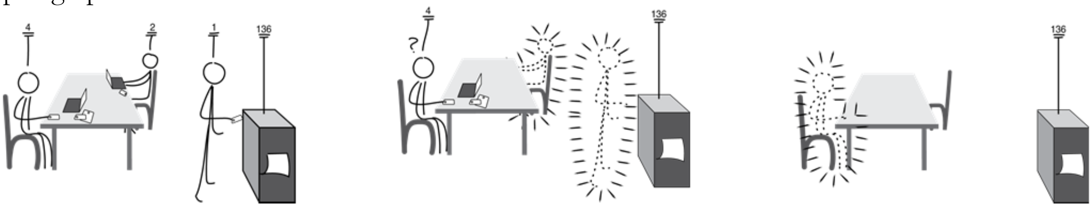


Figure 26: Detecting static devices by the number of macs and the days detected. Only the username connected with the printers is left.

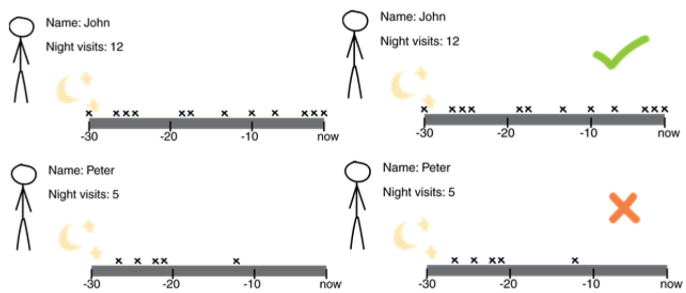


Figure 27: Detecting users who connect to the Eduroam network after closing times for multiple days

5.2. Visits between SLs in relation to traffic

To determine the occupation at the aggregated levels SL0-2, the direct use of the session data in the *wifilog* is insufficient. The raw session data does not show whether a session of five minutes is one created by a pedestrian strolling by or a student standing in line at the faculty canteen. With just session data, there is no difference between traffic and non-traffic sessions. By creating visits, the distinction between traffic and non-traffic can be made.

The term traffic has different interpretations on different SLs. For SL0 and SL1 traffic consists of cyclists, pedestrians and other users on the move outside educational facilities. For SL2, the term includes the traffic as defined for SL0 and SL1, but also users moving between floors.

Outdoor traffic can be detected on SL0 and SL1. On these SLs, a visit is defined as the total consecutive sessions a person is at a single educational facility. If a user leaves the facility and is detected at a new facility, a new session is started. If the user then decides to move back to the first educational facility, the data is not grouped to the first visit at that facility, instead a new visit is created (Figure 28). Outdoor traffic is often destination based, so users in transit will have short visits to facilities where they happen to pass by until they have reached their destination, where they will stay for a longer time. To remove outdoor traffic, short visits to faculties need to be removed.

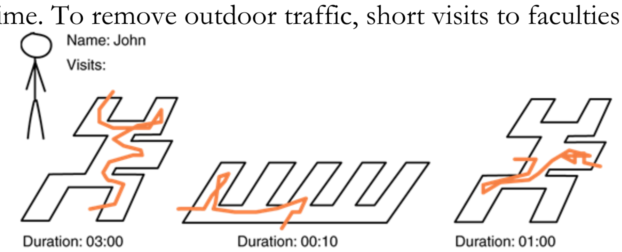


Figure 28: Multiple Visits for a Single User.

On SL2, a visit is the total time of consecutive sessions spent on the same floor in the same educational facility. Visits cannot be used to detect outdoor traffic, because the average visit time is lower compared to SL0 and SL1. To remove outdoor traffic, the SL2 visits have to be cross-referenced to the visits from SL1. To remove traffic between floors the visit duration can simply be limited to be larger than a pre-set quantity.

5.3. Filtering methods and grouping

The filtering and grouping of data is performed in multiple queries, which all output either a table or a materialized view. Queries are discussed in order of creation, the code for the query can be found in Appendix VI under the same name as used in this paragraph. All queries are accompanied with a flowchart depicting the steps made in the queries, the input tables and the output table, as well as a figure showing the column names and an example record of the output.

5.3.1. Query 1: Identifying Static Device Users

Static devices make up around 0.004% of the users and 0.05% of the sessions in the *wifilog* for April. However, the session times are long, amplifying their effect on the final occupation numbers. Before removing the static devices, they first have to be identified. The identification is done using the last thirty days of data in the *wifilog* table (Figure 28). The threshold for the maximum number of macs is based on the amount of different connected devices an person in the Netherlands holds, casted to a round number (Statista, 2014). The derived maximum of four is used to list possible static devices. For each of the resulting usernames, the average number of devices for the last thirty days is counted. If this number exceeds the average of three devices, the connected username is considered to be connected to static devices, as human behaviour is generally linked with variable behaviour, rather than static occupation. The output is a materialized view named g01_static_devices. The columns for the output and the first row are shown in table 15.

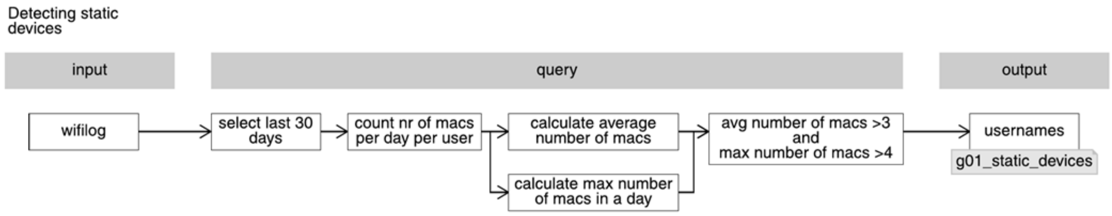


Figure 29: Flowchart of Query 1.

	username text	avg_macs numeric	max_macs bigint
1	4egdGj0KUy9s1Cb83eaqd/igsWM/MfpI60xZLr0tay4=	75	96

Table 15: g01_static_devices Columns and an Example Record.

5.3.2. Query 2-4: Identifying Users Living on Campus

Queries 2, 3 and 4 are used to identify users living on or near the TU Delft campus. These users might have the possibility to access the Eduroam network from their home, leading to these users to be registered at an educational facility while they are not in the building. The results from the three queries enable the removal of sessions of specific users on specific APs instead of removing complete users from the data. As this process only removes AP specific records, it needs to be executed before any aggregation processes.

Query 2 and query 3 are intermediate steps for the identification of users living on campus. Query 2 detects users frequently on campus after closing time, query 3 finds the most used AP of users who have been detected at night. The main difference between the two queries lays in that the latter finds the most used AP for each user detected at night for the last thirty days while the former finds which user is detected more than ten out of thirty days. Query 4 uses the result from the two queries as input and delivers *usernames* and favourite AP for users living on campus. All three queries are shown in figure 30, a result of query 4 is shown in table 16.

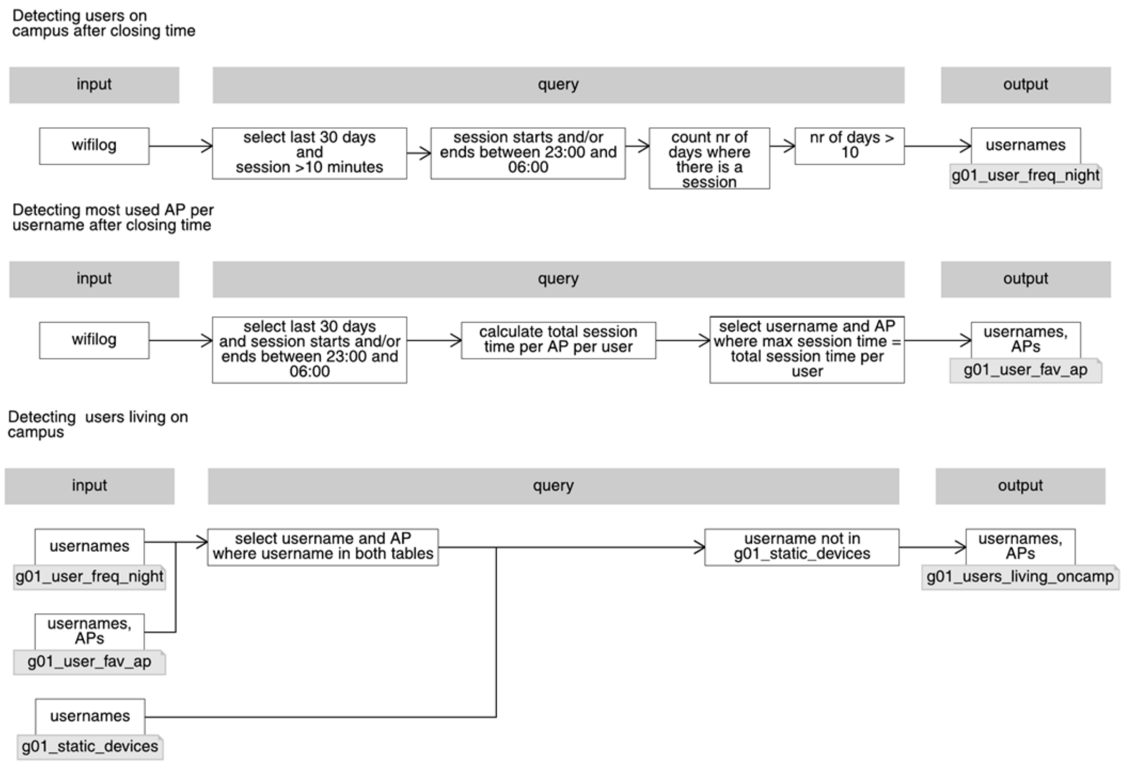


Figure 30: Flowcharts of query 2-4

	username text	apname text
1	0gzpA2mUhU1cu3GSAmpJSv506Lfhw81WERH9NIcWGAA=	A-23-0-046

Table 16: Results from g01_users_living_oncamp.

5.3.3. Query 5: Visit Creation and Filtering of Users Living on Campus for SL1

The main purpose of query 5 is to aggregate sessions into visits. Data is loaded into the *wifilog* per day, after which the concerned date is stored in the table g01_processed. The query only selects data from the *wifilog* table that has yet to be processed. By limiting the selection to unprocessed data the execution time is shorter, the temporary space used by PostgreSQL is smaller and the resulting materialized view is smaller.

To create the visits for SL1, changes in building number are monitored by comparing the building number of the current record with the building number of the last record. When a change occurs, a Boolean is returned, used to sum consecutive sessions marked to be inside the same building. The resulting visits are stored in the materialized view *g01_sl1_vnol*. VNOL stands for Visits No users Living on campus. The query flowchart is shown in figure 31, and the results in table 17.

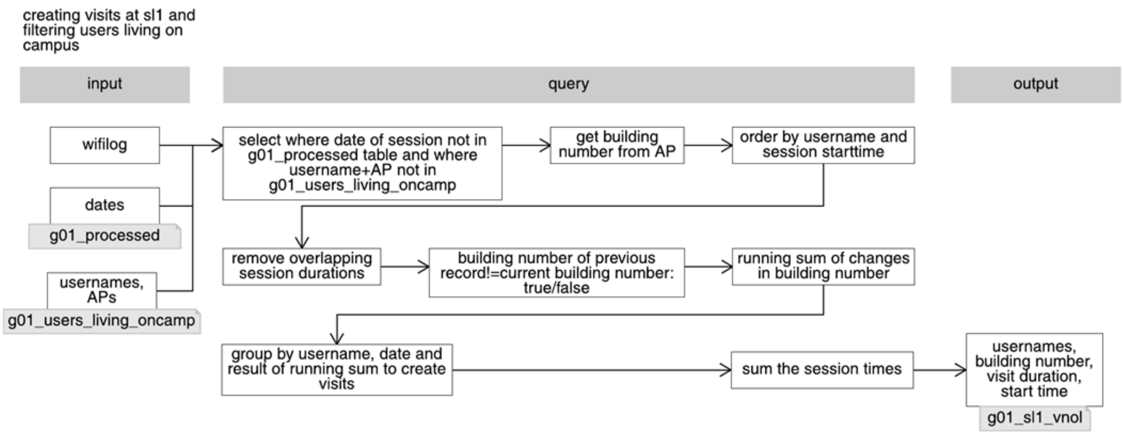


Figure 31: Flowchart of Query 5.

	username text	building_no integer	fac_visdur interval	start_t timestamp without time zone
1	vqGOfBIg0t3Q+ZsnV5wHuvTjbnVRKZq65T/sabPlc1I=	62	00:15:13	2016-04-25 11:03:10

Table 17: Columns and a Record from g01_sl1_vnol.

5.3.4. Query 6: Calculating Occupation and Exploitation for SL1

The resulting table from query 5 is used as the main input to calculate the occupation and exploitation per building, per day, per hour. First, static device visits are deleted by cross-referencing to the results of query 1 (Figure 32). Outdoor traffic visits are removed by filtering out all visits shorter than fifteen minutes. The output can be used to compute occupation per facility, per user, per hour. The UFA stored in the database is used to calculate exploitation. The table containing the UFA data is also used to filter out buildings that are not in the scope of the project. The resulting records contain the building id, a date, the hour, the occupation and exploitation (Table 18).

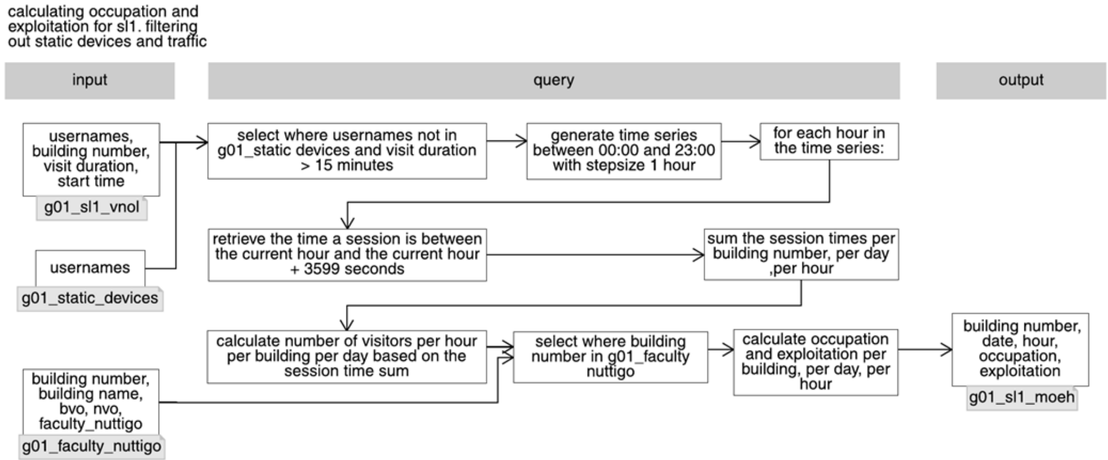


Figure 32: Flowchart of query 6

	buildingid integer	datum date	l_hour time without time zone	occupation double precision	exploitation double precision
1	3	2016-04-01	00:00:00	0	6983

Table 18: Columns and a Record from g01_sl1_moeh

5.3.5. Query 7: Calculating occupation and exploitation for SL0

The calculation of occupation and exploitation on SL0 predominantly uses the same steps used to perform the calculation for SL1. However, the difference is in the grouping; for SL1 the calculation is performed per building, per day, per hour, while the calculation for SL0 is performed per day per hour (Figure 33). As with SL1, the UFA per educational facility is used to move from occupation to exploitation, but because of the change in grouping, the UFA has to be summed into the total campus UFA. The resulting records are stored in the *g01_sl0_moeh* view as shown in table 19. Here, MOEH stands for Materialized view Occupation and Exploitation per Hour.

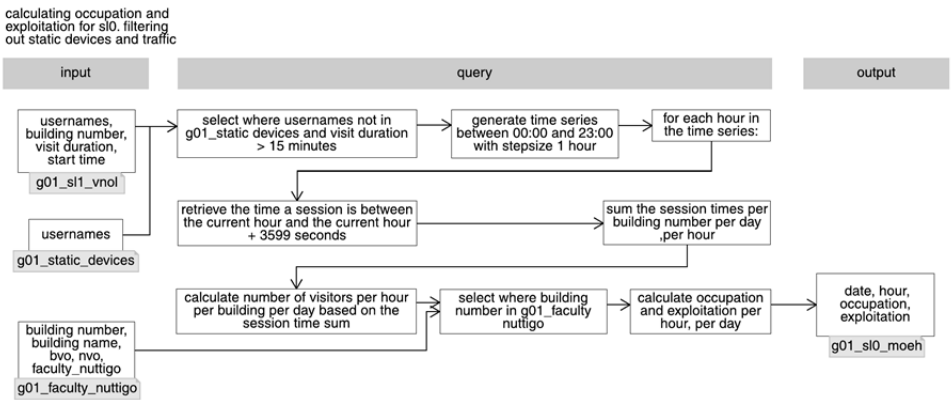


Figure 33: Flowchart of query 7

	datum date	l_hour time without time zone	occupation double precision	exploitation double precision
1	2016-04-01	00:00:00	163	1489

Table 19: Columns and a Record from g01_sl0_moeh

5.3.6. Query 8: Visit Creation and Filtering of Users Living on Campus for SL2

SL2 requires a different handling of visits, as they should be grouped by floor instead of building. Furthermore, the threshold for traffic is between floors is set at 5 minutes. The creation of SL2 visits is similar to the SL1 query, except the running sum is based on subsequent records on the same floor. The query is depicted in figure 34 and the resulting materialized view in table 20.

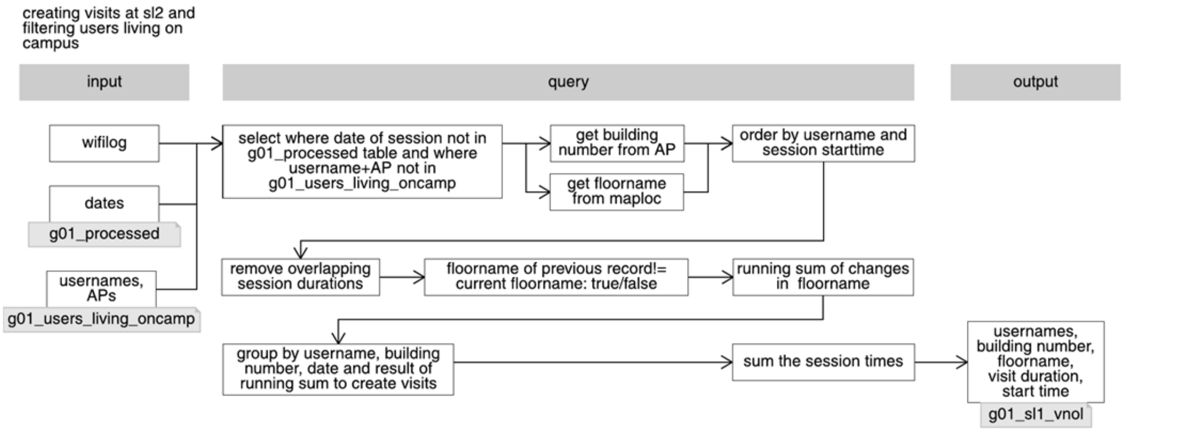


Figure 34: Flowchart of query 8

	username text	building_no integer	floor text	floor_visdur interval	start_t timestamp without time zone
1	dgr7MXOzhNG0jgay1YwAC103qRyF59736jI6d31tzLQ=	8	begane grond	00:05:09	2016-04-07 10:04:39

Table 20: Columns and a Record from g01_sl2_vnol

5.3.7. Query 9: Calculating occupation and exploitation for SL2

As mentioned in paragraph 5.2, SL2 visits cannot directly be used to filter outdoor traffic, only users moving between floors. To recognise floor visits that are part of outdoor traffic, SL2 visits are cross-referenced with visits from SL1 shorter than 15 minutes, and removing visits which match on username, building number and start within the time of the SL1 visit (Figure 35). After removing outside traffic, the users moving between floors are removed by filtering the visits shorter than 5 minutes. To calculate the exploitation, the UFA per floor from the *g01_floor_nuttigo* is used. The resulting materialized view is structured similarly to the results from the occupation and exploitation calculation on SL1 (Query 6), however for SL 2 the floor name is added to the view (Table 21).

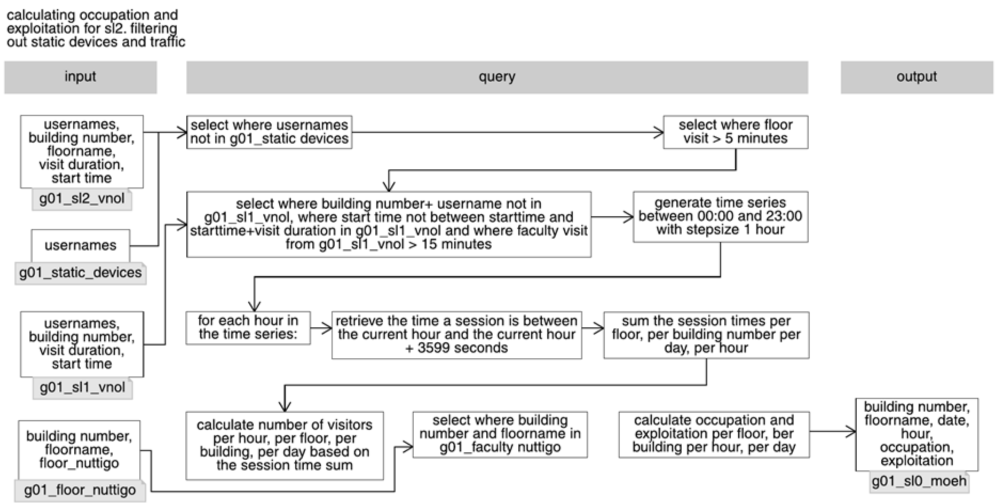


Figure 35: Flowchart of query 9

	buildingid integer	floor text	datum date	l_hour time without time zone	occupation double precision	exploitation double precision
1	3	kelder	2016-05-09	01:00:00	0	11

Table 21: Columns and a Record from g01_sl2_moeh

5.4. Effects of Filtering and Grouping on Data Flow

For each query in paragraph 5.3, either data is discarded or aggregated. Every query discussed has a different effect, as well as having the possibility to contain both code used to filter data and code used to group data. The effect of each grouping and filtering step is shown in figure 32. The effect of grouping activities is portrayed in ratios while filtering effects are portrayed in percentages. For each step, the number of sessions, visits or records is documented. The *wifilog* in figure 32 is limited to all session recorded in April to reduce the load times between queries.

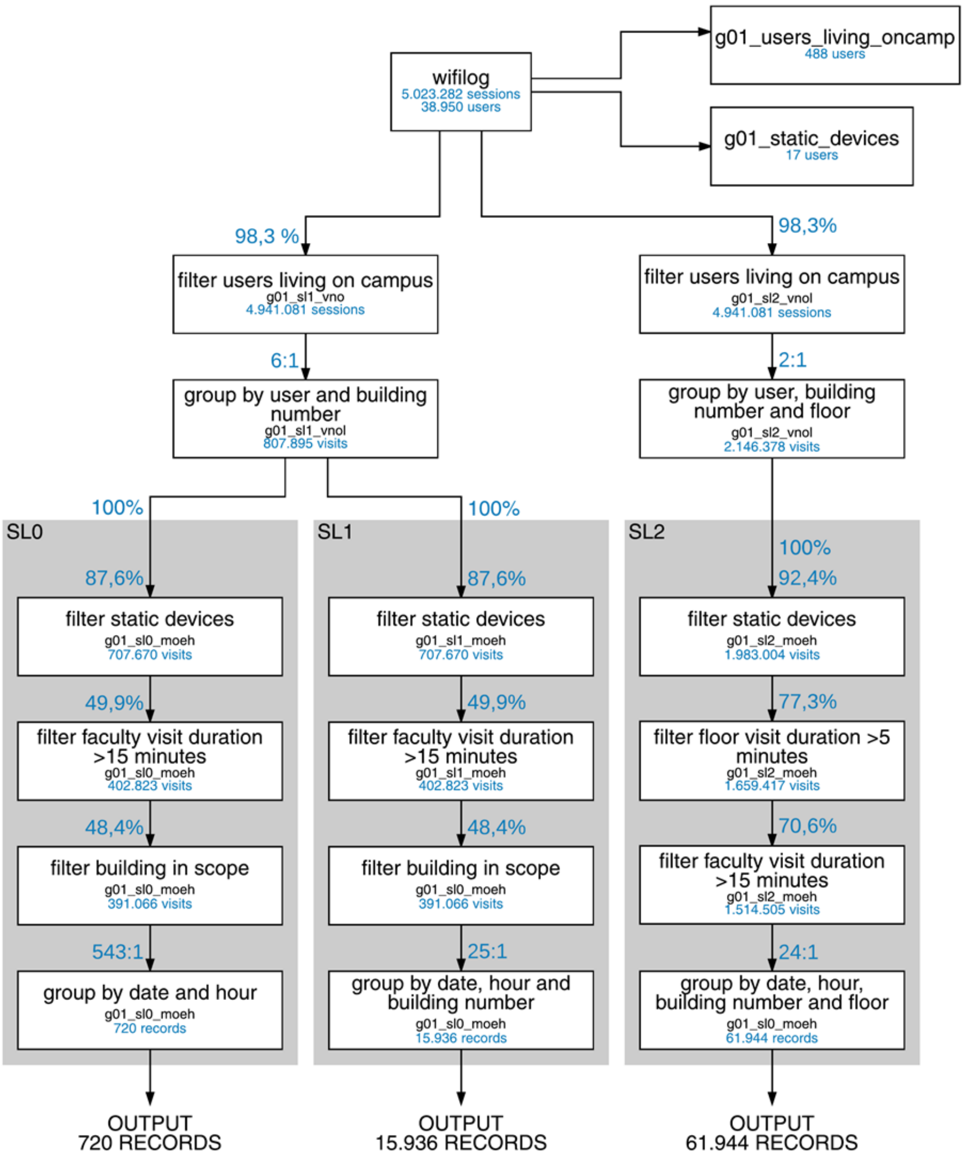


Figure 36: The effect of filtering and grouping on the data stream.

Figure 36 shows that the 488 people who live on the campus, 0.01% of the users in the *wifilog* table, are responsible for 1.7% of the sessions made in April. Another observation is that 37.7% of the visits on SL1 are classified as outdoor traffic, while only 6.3% of visits on SL2 are classified as outdoor traffic. There are two possible causes; either many of the visits on SL2 which are part of outdoor traffic are already filtered out in the previous step, or a SL1 visit classified as outdoor traffic consists of less SL2 visits than an SL1 visits which is not classified as outdoor traffic.

6. Data Storage

The focus of this chapter is on the storage of the data and information generated from and during the data analysis. The topics discussed in this chapter include the table composition, view dependencies, the order of creation for tables and the update structure of the materialized views and tables.

6.1. Database Structure

After processing, the database consists of a series of interdependent tables, all representing a step in the analysing and aggregating process. As the database provides storage for all Synthesis Project group outputs, each table name starts with g01, as to indicate this is a project in_sight table.

6.1.1. Table composition

The results of the queries discussed in chapter 5 are stored in materialized views. Figure 37 shows an overview of all these materialized views, their column names and column types.

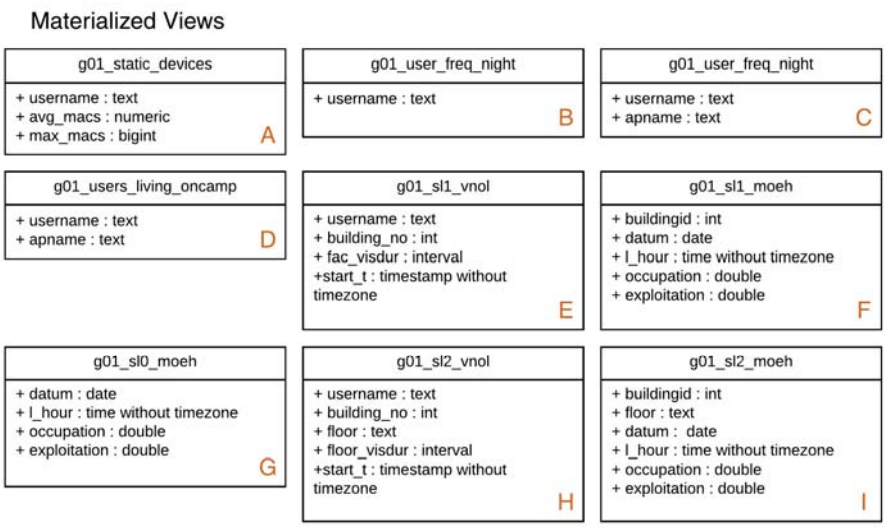


Figure 37: An overview of all materialized views with their respective columns

Many of the tables shown in figure 38 are not used in the data analysis stage. Tables containing geometry, such as *g01_sl2geom*, are needed for the communication stage. For example, *g01_sl2geom* contains the polygon geometry of all floors in the project scope as well as their height and the start height (*minHeight*) of the floor. The 38.2, 38.3 and 38.4 tables are used as the permanent storage place for the results generated in figure 37.F, 37.G and 37.I.

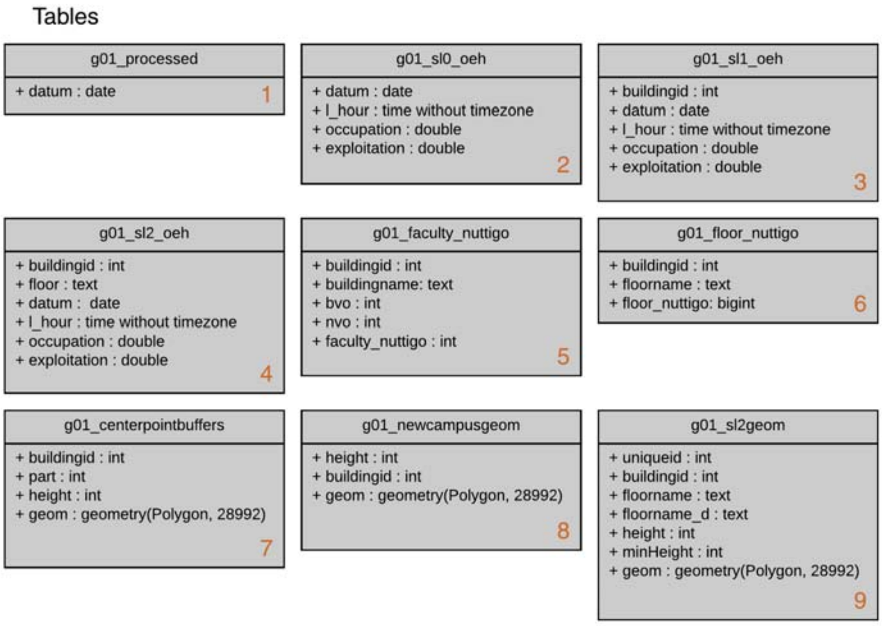


Figure 38: An overview of all tables with their respective columns

6.1.2. View Dependencies

The use of materialized views creates a network of dependencies. Materialized views can be refreshed, and the SQL query that created them is rerun to fill the materialized view with new data (The PostgreSQL Global Development Group, 2016a). Because the query is stored when creating a materialized view, every other view or table used in the query should exist at all times. Any table or other materialized view connected as input to a materialized view in figure 39 is a dependency.

The dependencies in the database limit the order of creation of materialised views, the order in which the queries are discussed in chapter 5.3 allows for all views to be created without having any conflicts of missing inputs, though before starting the tables used as input in figure 30 (*wifilog, g01_processed, g01_faculty_nuttigo and g01_floor_nuttigo*) should be available.

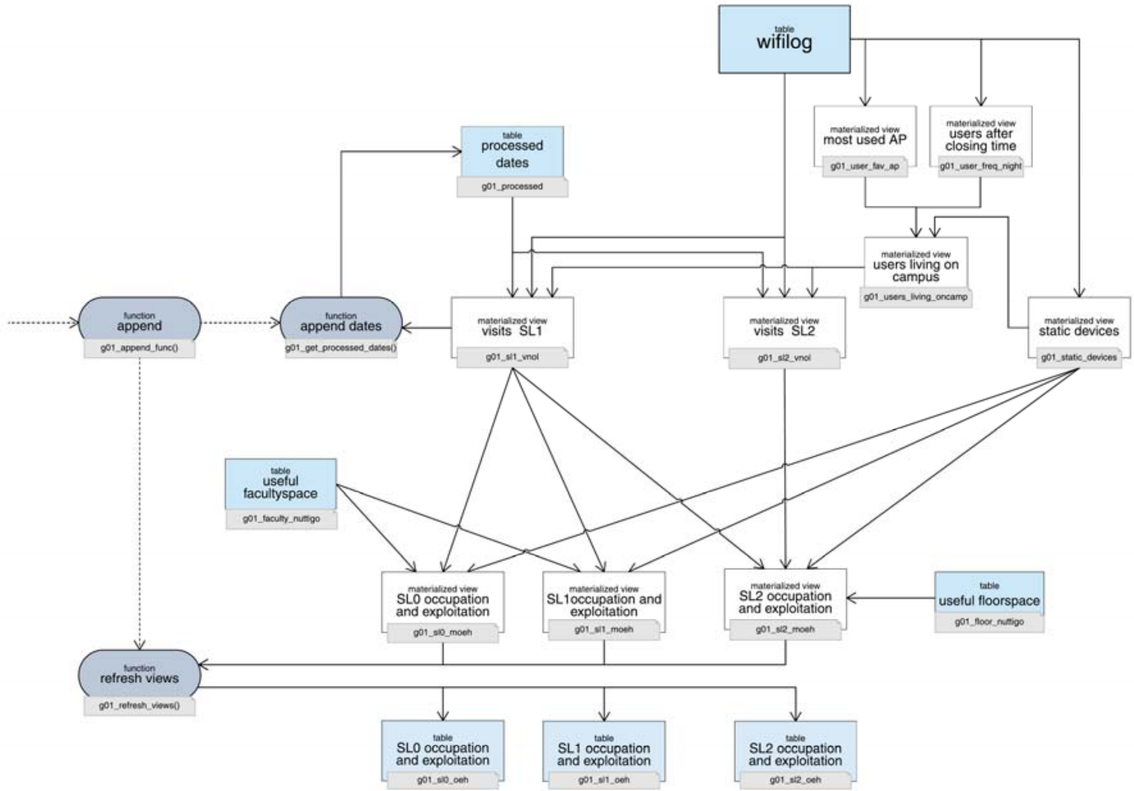


Figure 39: All Database Tables and their Relations.

6.2. Processing new data

The structure of tables, queries and materialized views is created with to focus on allowing new data added to the *wifilog* table to be processed. There are four advantages to automatically update the occupation and exploitation numbers based on new data from *wifilog*:

- The information provided to FMRE is always near historical
- The relevance of the dashboard continues to increase after the project is concluded.
- By combining the ability to process new data with an efficient query plan, only new data has to be processed instead of the complete *wifilog* table.
- New data does not have to be processed manually.

The process of analysing new data is controlled by the *g01_refresh_views()* function. This function is stored in a crontab, which is used to schedule background tasks on UNIX based systems (The Open Group, 2013). The crontab is run on an external computer under the supervision of Drs. C.W. Quak and is executed every day at 12 past 2 in the morning. The function refreshes all materialized views and executes two other functions: *g01_append_func()* and *g01_get_processed_dates()*. The first function appends the results from the materialized views in figures 28.F, 28.G and 28.I to tables that contain the final results for each SL (Figure 37). The second function retrieves all dates that have been processed in one run and appends them to the *g01_processed* table. This table in turn is used by a materialized view to select unprocessed data. The function to append the processed dates is executed after the append function to ensure that dates cannot be appended without appending occupation and exploitation numbers.

7. Data Communication

The practical output for the project is an online dashboard as an SDSS, on which FMRE can view the occupation and exploitation of available space. The dashboard is partially created for a subsequent course, Geoweb Technologies. The chosen subject focusses on the 3D visualisation of occupation data for SL2.

7.1. Boundary Conditions

The conditions for the dashboard are described in paragraph 3.2.3 and 3.2.4, as requirements specified by FMRE. Additionally, the 3D map should be linked to statistics published onto the dashboard. The dashboard should be web based and provide an active connection with the database. The separate tasks to be executed are defined and prioritized by using the MoSCoW approach and can be seen in table 22.

Must	Should
• SL0: o Full 3D Model (LOD1). o Interactive Graphs. • SL1: o Coloured Buildings (Exploitation). • Popup: o Building Name. o No. of Users. • Slider: o Choose Day. • Legend (Colours). • Mock-ups: o Coloured Buildings. o Generic Footprint, Different Heights. o Coloured per Floor.	• SL1: o Bar Chart Geometry (Exploitation). o Outlines for Ideal and Footprint. • SL2: o Coloured Floors. • Popup: o Reliability Figure. • Slider: o Choose Hour.
Could	**Won't**
• SL2: o Full Campus Application. • Slider: o Average Values for Several Days. • Export: o Export data as image. o Export data to csv. • Automatically Update Calendar Range	• SL2: o "Scroll" Through Floors. o Heat map (Exploitation). o On Click: Visible Floor Separation. • SL3-4: o Publication of Data. • Slider: o Average Values for Several Hours.

Table 22: MoSCoW Rules for Dashboard.

7.2. Dashboard

The following paragraphs are meant to explain what tasks the dashboard can perform. A very short technical description is added.

7.2.1. User Interface

In concept, the dashboard exists of three partitions: a menu and navigation area, a map area (Figure 40) and a hidden statistics area that can be opened on demand (visible in figures 40 and 41). The menu contains a slider that allows the user to specify which SL to show. Below, the user may select a date from the calendar for which the data will be loaded. For now, only April and May 2016 are available. A slider can be used to choose a time to display, in intervals of one hour. A value picker is placed to choose an ideal exploitation value, which determines the building classification. On the bottom of the menu, a legend is placed for the classification, using intuitive colours for classes 'calm' (blue), 'ideal' (green) and 'busy' (red). The map area allows panning, rotating and zooming.

Figure 40: Dashboard Facility Level, Realistic Geometry, Graphs for Aula Conference Centre.

Campus Level

When the dashboard opens, full campus geometry is visible, connected with information about SL0 (Figure 40). The map area shows realistic LOD1 geometry, all classified according to exploitation as an aggregated campus rate. The exploitation is taken specifically for the selected hour. When the graph icon on the right side is clicked, the statistics area pops up, showing occupation and exploitation as defined in chapter one, for the entire selected day.

Figure 41: Dashboard Campus Level.

Facility Level

When the level slider is switched to facility level, the realistic campus geometry is classified per building according to SL1 information (Figure 40). Underneath the slider two buttons appear, with which the geometry type can be switched between realistic and bar chart geometry (Figure 42). When a building is clicked, the graphs in the statistics area show information about the selected building, and are updated when a different building is clicked. Again, the graphs display the entire selected day and the colours represent the selected hour.

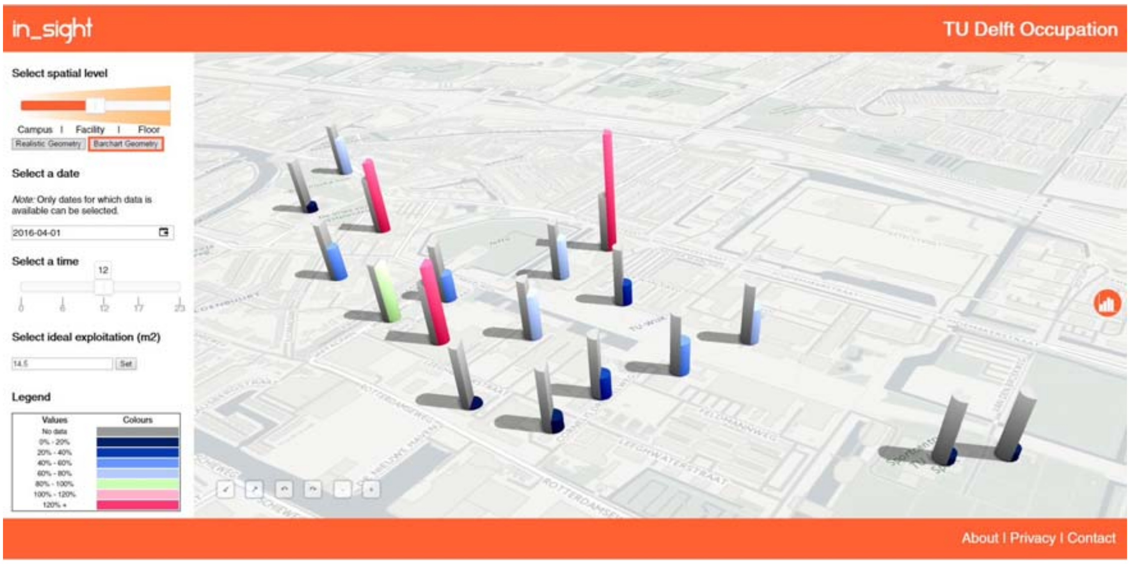


Figure 42: Dashboard Facility Level, showing bar chart geometry

Floor Level

The third option on the level slider is SL2, where the full campus geometry is split into floor levels (Figure 43). Each floor is classified according to the selected date and time. When a floor is selected, the graphs display its data for the selected day. It must be noted that any output for this level is not reliable due to the way of measuring the location of users (paragraph 8.1).

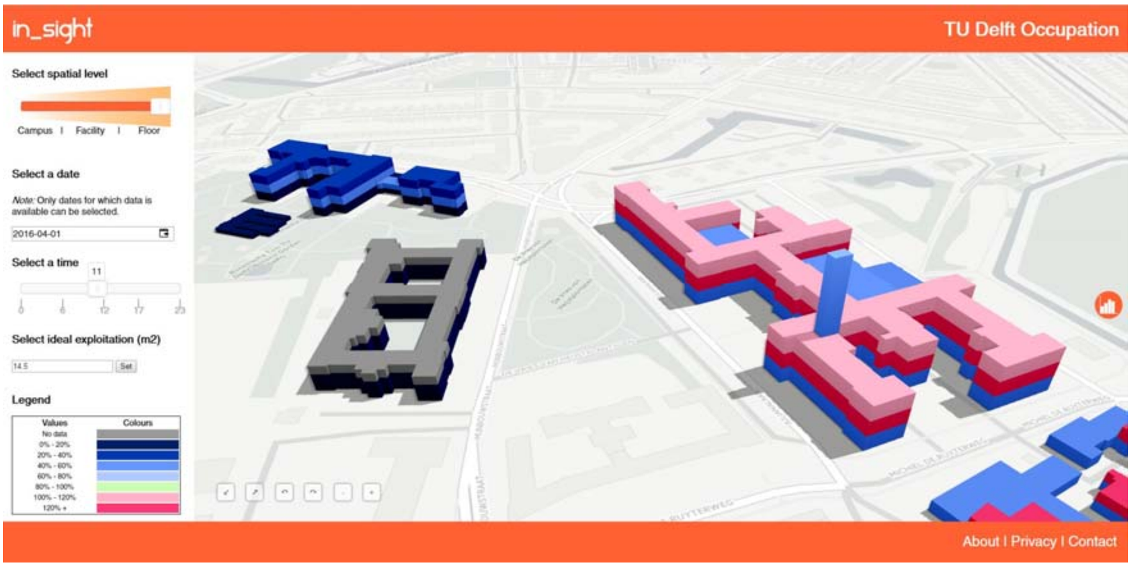


Figure 43: Dashboard Floor Level, no Chart.

Information Pages

Three extra pages have been added to the dashboard, to provide some additional information about the goal of the dashboard and contact information of the team. Altogether, they provide information required when publishing personal data, according to the European Data Protection Directive. These pages may be accessed from the bottom bar. The in_sight logo in the upper left corner of the page can always be used to refresh the page, or return to the home page.

7.2.2. Technical Description

The back-end communication of the dashboard is depicted in figure 44. All data sources mentioned in paragraph 2.4.1. are stored into the PostgreSQL database. The host software, GeoServer, utilizes encoded WFS requests to retrieve data from the database, according to the SL, geometry type and day that is requested on the dashboard. The host translates all requested data into valid GeoJSON objects, which are returned to the client. The client runs on WebGL with an OSM Buildings viewer implemented, which together produce the classified buildings. The viewer transforms the retrieved 2D geometry and attributes into classified 3D blocks on the fly. The classified geometry is projected onto a background layer retrieved from CartoDB. The sliders, calendar and charts depend on embedded open source libraries. Further technical details can be found in the Geoweb Dashboard Documentation.

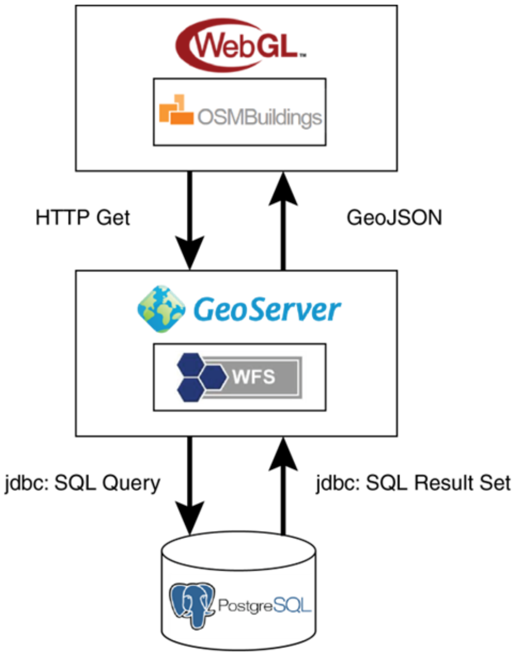


Figure 44: Back-end Communication Dashboard.

7.3. Progress

Currently, the dashboard is functional and performs most of the required tasks. However, certain parts may still be improved:

- Sometimes when the dashboard is loaded, no geometry shows up. It appears as though this problem is caused by a connection deficiency, though it is not clear if this is depended on the background layer or the database.
- The sliders for SL and time of day work, as long as a new layer is loaded by a single click. When a different layer is requested too soon, the dashboard will paste several geometries on top of each other.
- Currently, no functional popup has been added. A setup for this is available in the code, but the implementation has not been successful so far. This is probably due to the fact that OSM Buildings on WebGL is still in beta version.
- The database should be capable of handling the publication of real-time data. However, as the database is not updated more often than once every day, a choice has been made to only publish data from April and May 2016.

8. Data Quality

All through the methodology process, as described in paragraph 3.5, quality control is an important and overarching subject that has to be taken into account. For each part of the process, the term quality may refer to different influential aspects, like representativeness, reliability, correctness, accuracy and legibility. The quality control process has been executed through theoretical research, the design of multiple case studies and the gathering of validation data. Throughout the process, the quality is also measured through a Key Performance Indicator (KPI) framework (Appendix VIII).

8.1. Data Capture

The representativeness of captured data can be checked through the execution of case studies. This paragraph describes four case studies to define the representativeness of data on SL2 and three case studies to test the possibility of drilling down to SL4.

8.1.1. AP Visibility Across Floors (SL2)

In order to test how well discernible different APs are across floors, a case study was designed. The concept was to record all visible APs on a few subsequent floors, as to mirror the AP tags stored in the database to reality.

Case Study 1: EWI Floor 4-6

The APs at the main building of EWI are grouped per three floors (Table 23). If spread evenly, each of these floors should contain two APs. To test discernibility, the measurements are taken across two floor groups. Altogether no more than twelve APs should be visible.

EWI Main Building	
18th-19th-20th Floor	**21st-22nd Floor**
A-36-0-018	A-36-0-021
A-36-0-019	A-36-0-043
A-36-0-020	
A-36-0-042	
A-36-0-071	
A-36-0-072	

Table 23: Possibly Available APs on EWI Floor 20-22.

The case study returns an output of twenty four different APs on the three tested floors, twice as many as the proposed maximum. A third of these extra APs, id 16 and up, show up on the fourth floor only. An explanation may be that these APs belong to the annex, which spans over four floors and is directly connected to the main building via skyways. The remaining fifteen however, still represent more APs than should be available. Furthermore, two APs are highlighted in figure 45.A that are visible over the span of all three floors, with an RSSI so similar, that it cannot be decided from these measurements, to even which group of floors they belong.

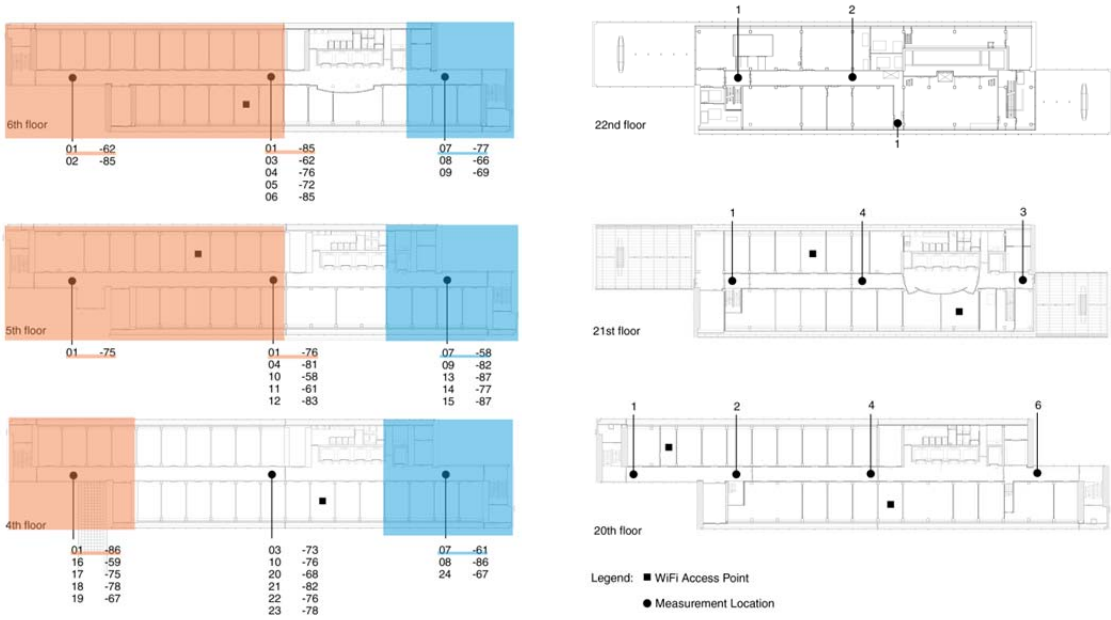


Figure 45: A. AP Visibility on EWI Floor 4-6. B. Number of APs visible on EWI Floor 20-22.

Case Study 2: EWI Floor 20-22

In order to test the actual influence of the annex, the AP visibility case study was repeated at the top floors of the EWI building (Figure 45.B). The 21st and 22nd floor should contain two APs, floor 18-20 should contain six. With a maximum of eight visible APs, the six different recorded ones prove a higher reliability of the *wifilog*. , it remains interesting to see that the north side of the building, right on the pictures, passes more WiFi signals than the rest of the building.

Conclusion

An AP, at least in the EWI building, may be visible across several floors and even several building partitions. Therefore, when retrieving data from the database according to SL2, the output probably contains a rather low reliability.

8.1.2. Case Study: AP Connectivity Across Floors (SL2)

To expand the visibility test, a similar experiment was designed, with a twofold goal. First, to test if the AP emitting the strongest signal at a certain point is actually on the floor the measurement was taken at. Second, the test is taken in two different buildings, to test the influence of floors and walls on signal propagation. The case study is executed as a series of close range measurements. At each measurement location, the used device is disconnected, and then reconnected to the Eduroam network. The strongest signal is stored with the APs mac address. According to the signal strength and visual inquiry, an AP map can be made for the tested areas.

Case Study 3: EWI Floor 20-22.

A first case study area is the top three floors of the EWI building, as a set of isolated floors within a building with a steel frame. First, all possibly visible APs are listed (Table 23). From the taken close range measurements, the actual locations of APs and their mac addresses are rather easy to discern, so that the case study measurements can be connected to APs stored in the database.

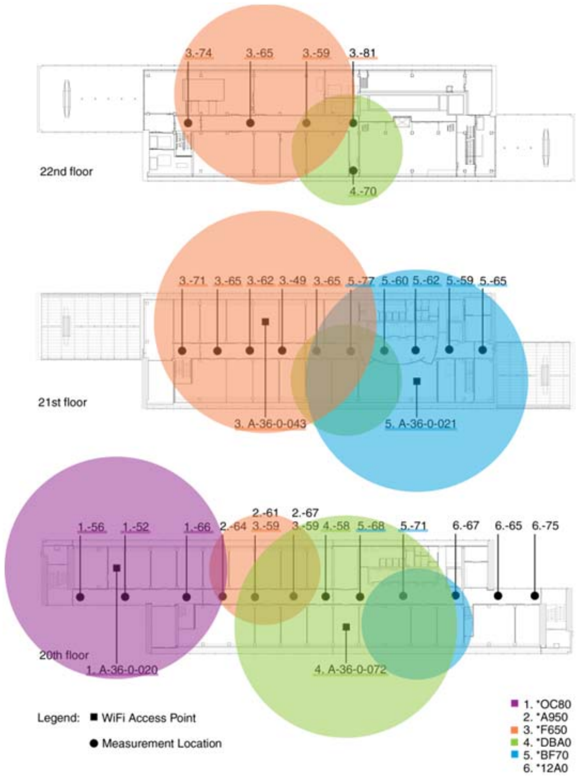


Figure 46: AP Connectivity on EWI Floor 20-22.

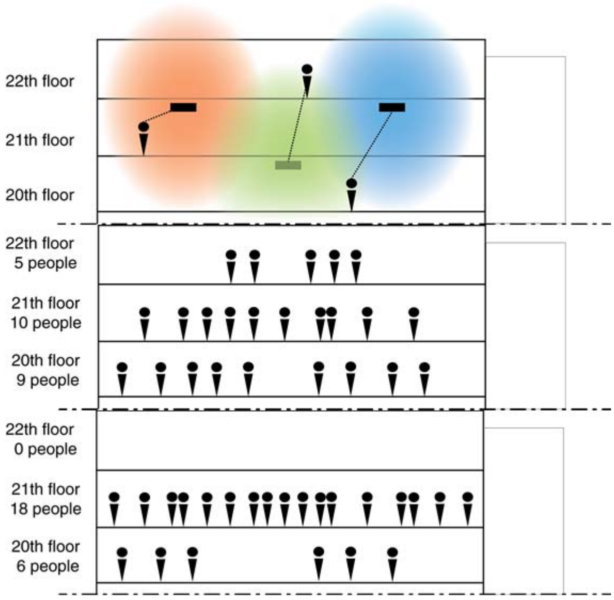


Figure 47: Effects of WiFi System Settings on SL2, Directly Derived from Case Study.

The results of the case study are depicted in figure 46. As expected from earlier case studies, it is not only difficult to discern on which floor a connected device could be, but also on which group of floors. The AP with id 3 was the strongest connecting AP for certain locations on all three tested floors. The data retrieved from AP 1 might be reliable according to this test case, but its signal may as well reach across lower floors. On the 20th floor, where the signals of APs 4 and 5 are rather similar, the testing device connected with AP 5 several times, even when in very close range of AP 4. Most striking is the connection to AP 4 on the 22nd floor, where it was never even recorded on the 21st floor. Therefore, the reliability of gathered data at SL2 will be very low, at least for a steel skeleton building like EWI. The direct effects on the data can be seen in figure 47. The incorrect placement of users makes it appear as though no users are present on the 22th floor, and the 21st floor is more crowded than it should be. It must be noted that the data for AP with id 1 has been removed for this visualization. The case study result implies that a middle floor may always be marked more crowded than it is, as more users have a chance of being connected to a middle floor AP.

Case Study 4: Architecture North-West Wing

To test the influence of a buildings' construction on WiFi signal propagation and the reliability of the former case study, the experiment was repeated at a rather isolated wing in the faculty of Architecture. This building contains thick, solid walls and floors, in contrast to the thin skeleton of the EWI building. However, a different type of APs is used, capable of emitting a stronger signal. Again, first the possibly available APs were listed (Table 24).

Architecture North-West Wing (F)				
Ground Floor	Ground Floor +	1st Floor	1st Floor +	2nd Floor
A-08-F-001	A-08-F-004	A-08-F-101	A-08-F-106	A-08-F-201
A-08-F-002	A-08-F-005	A-08-F-102	A-08-F-107	A-08-F-202
A-08-F-003		A-08-F-103		
		A-08-F-105		
		A-08-F-108		

Table 24: Available APs in Architecture North-West Wing.

The test was again mapped according to the measured area, and possible range for a select number of APs (Figure 48). It must be noted that both + floors only represent sublevels, and therefore only provide additional floors and walls, but not additional space between the three measured subsequent floors.

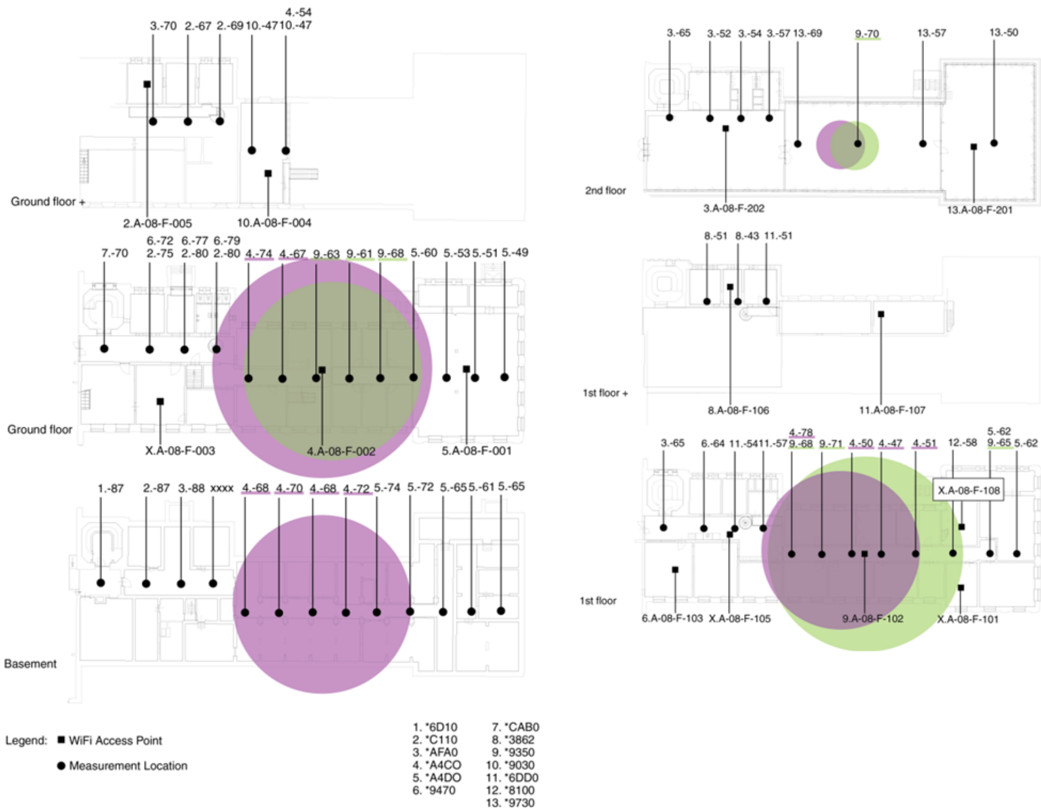


Figure 48: AP Connectivity in Architecture North West Wing, AP 4 and 9.

An interesting event in the case study is the interchangeable connection with APs 4 and 9 (Figure 48). Though they are placed on different floors, the ground and first respectively, the received signal strengths from these APs are largely similar. In one case, a connection was made with AP 9 one floor above, while standing right underneath AP4.

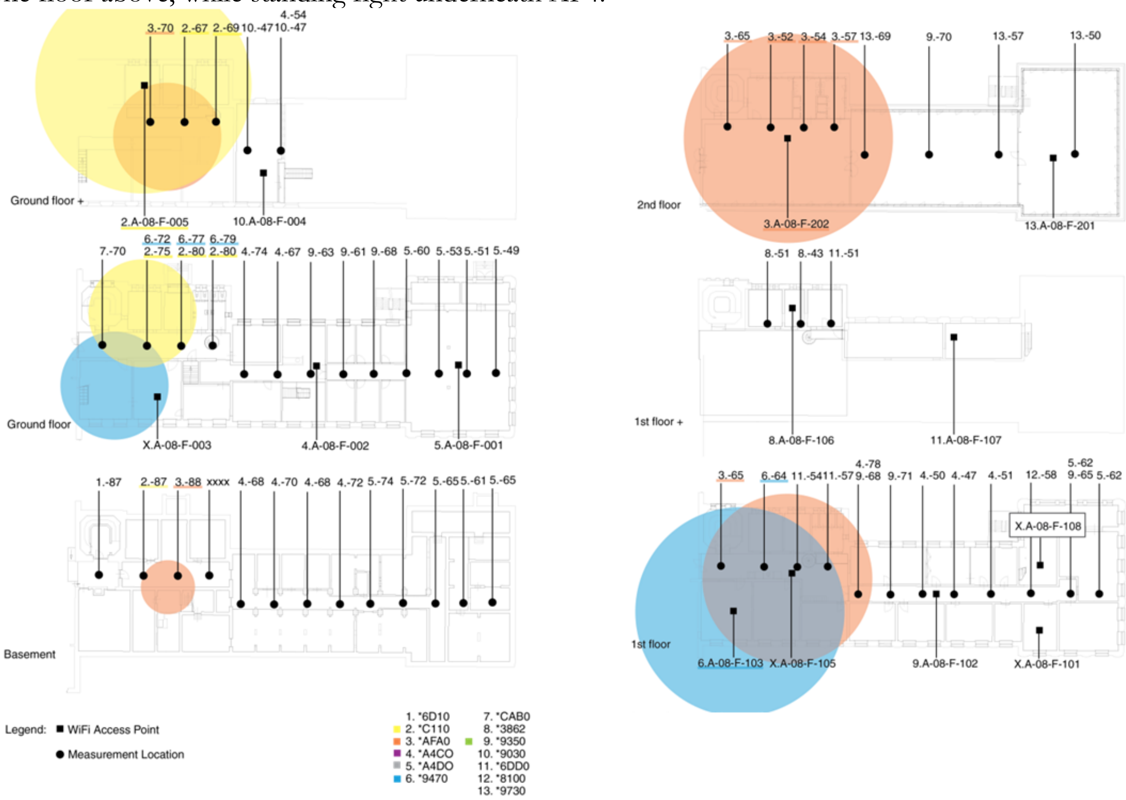


Figure 49: AP Connectivity in Architecture North West Wing, AP 2, 3 and 6.

The expectation for this case study was a clearer discernibility between floors, as the construction of the building is much heavier than that of the EWI building and each floor is at least twice is high. However, the received signals on each measurement location are so strong, that the main conclusion to be taken from this test is that the AP system is sufficiently placed to provide the entire wing with WiFi, even if one or several APs fail. Even the largest part of the basement is covered, which is used only for storing. Not even the air in this area is cleared, but the WiFi signal is generally stronger than in some parts of the second floor, where APs are available. The most striking basement connection is that with AP 3 (Figure 49), which is located on the second floor. The signal encounters APs 2 and 6 as competitors on several measurement locations, and which AP is chosen as connecting one, seems more random than logical.

Conclusion

The available AP system is set up to always provide a WiFi connection when inside a TU Delft facility. Since this active WiFi availability is achieved through a dense network of APs, it can hardly ever be predicted to which AP a device will connect. Consequently, a device cannot precisely be located. An approximate horizontal location may be possible, but no matter the construction of a building, the floor on which an AP is within the building it is recorded in, cannot accurately be derived. Furthermore, all middle floors will be marked busier than reality. This statement is proven by comparing output on the dashboard, an example can be found in figure 34. Therefore, all occupation and exploitation numbers to be derived on SL2 will contain a very low reliability.

8.1.3. Case Study: AP Range (SL4)

So far, the visibility of an AP has only been visualized as a perfect circle with an estimated ratio, as to approximate global signal propagation. However, actual spread of a WiFi signal should not happen as smoothly and evenly. In order to actively test the before mentioned visibility of APs across floors and to map actual signal propagation, a case study focussed on a single AP was designed.

Case Study Background

Though a single AP may be recorded as belonging to a single room, the users recorded for that single room may not actually be inside. In order to make an easy step towards SL4, ideally each room would contain its own AP and record only users within that room (Figure 50). However, a user may be in between two APs, in a room not linked to either AP he could connect to. A second situation may be that there is only one AP for the user to connect to, but he will be recorded inside the room when he is not. In a third situation, a user may even be outside of the building, and still connected to an AP inside. Signal strength may be an indicator to filter these users out, though this cannot be done before clearly mapping signal propagation.

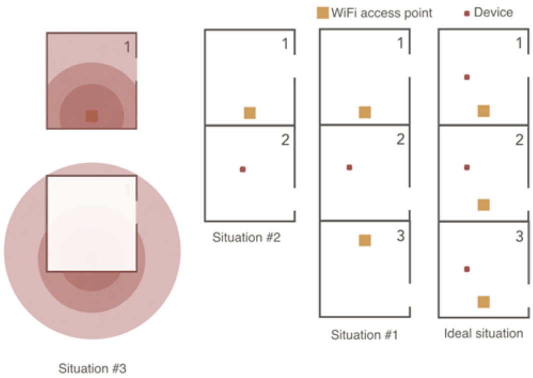


Figure 50: Situations Where a User may be recorded in a Room Different from the Real Location.

Case Study 5: Architecture Room T

As the faculty of Architecture is one of the buildings containing a heavier structure, a single AP within one of the rooms of this building was chosen for measuring signal propagation. A series of measurement locations was chosen to see how well and how evenly the signal would spread (Figure 51). The chosen AP is located in Room T, in the East Wing of Architecture.

Especially the measurements around the AP itself show a clear influence of the walls. Outside the room, a signal strength of -60 dBm is never reached. However, further spread of the signal appears rather unevenly distributed. In some cases, a close range measurement may even return a much lower value than one further away, though maybe less obstructed.

Conclusion

Signal propagation of WiFi APs may always be prone to similar influences, but its exact contours cannot easily be predicted. Signal strength has been proposed as a method to distinguish on which floor a user is. However, the test cases have proven that users are hardly discernible across floors. As signal still spreads vertically in a rather efficient way, SL2 cannot be determined with high reliability validity.

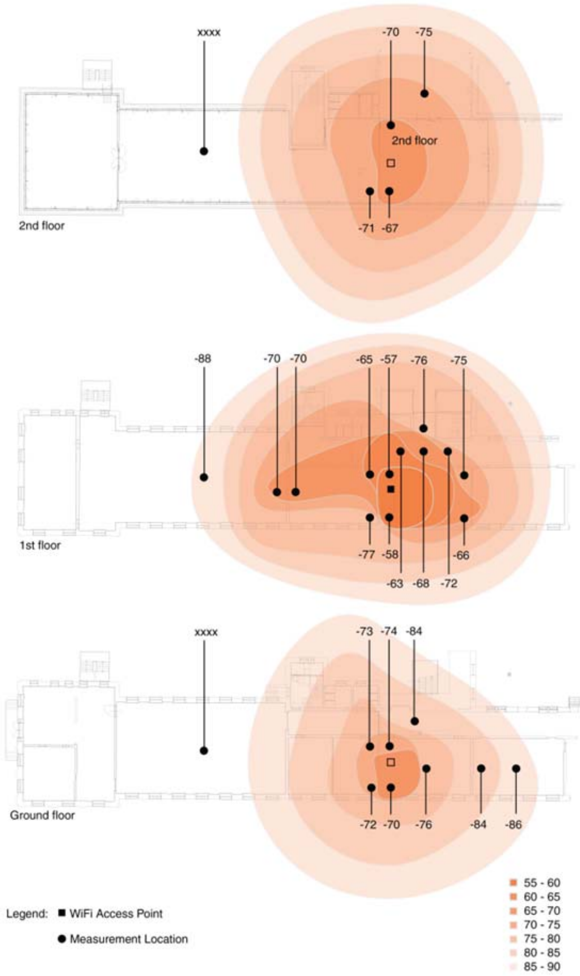


Figure 51: Signal Propagation of a Single AP, in Architecture Room T.

8.1.4. Case Study: Single AP, Multiple Rooms (SL4)

A first case study was performed in the faculty of Architecture, rooms B, P and Q. All users within these rooms should be recorded on a single AP, located in Room Q (Figure 52).

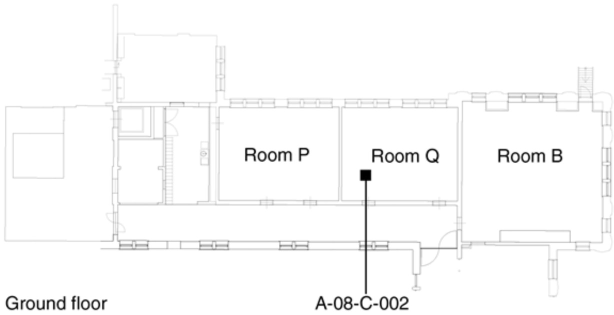


Figure 52: Architecture Rooms P, Q and B.

Case Study 6: Architecture Room Q per Hour

In order to test the representativeness of the gathered data, the distinct users around a single AP are counted along the course of a single day, with a temporal resolution of an hour. The counted values can be found in table 25.

26-apr		8	9	10	11	12	13	14	15	16	17	18	19
Room B	Schedule		x	x	x	x		x	x	x	x		
	Count		48	48	27	27	6	25	28	43	43	7	
Room Q	Schedule		x	x	x	x	x	x	x	x	x		
	Count		19	19	19	19	14	16	14	9	3	3	
Room P	Schedule		x	x	x	x	x	x	x	x	x		
	Count		18	18	18	18	10	10	10	10	0	0	
	Total		85	85	64	64	30	51	52	62	46	10	

Table 25: Ground Truth for April 26th at Architecture, Rooms B, P and Q.

The outcome is compared to both the raw *wifilog* data and the filtered data (Figure 45). As can be seen, the general trend is rather representative for the actual numbers. However, due to the temporal resolution chosen, sudden changes in occupation are lost. Especially change during lunchtime is not visible in the data. Furthermore, when a sudden change in occupation occurs, for example due to the end of a lecture, this is recorded later than its actual occurrence. For example, table 26 shows a drop in occupation between 10 and 11, after the end of a lecture. However, according to the data, the amount of people present drops an hour later (Figure 53). Beside this delay, it cannot be concluded from the data where exactly this drop in occupation occurs. From the ground truth it is known that people left Room B at this time, and that the occupation remains the same in both Room P and Q.

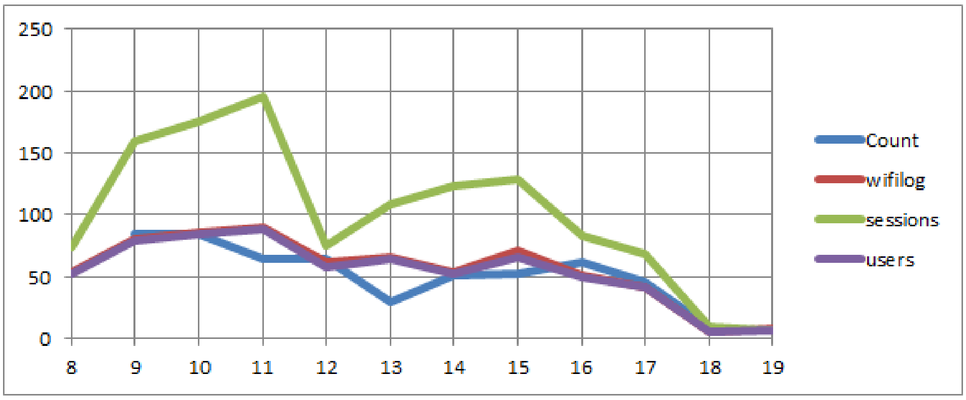


Figure 53: The Actual Amount of People Compared to Retrieved Data at Architecture, Rooms B, P and Q.

Another concept proven by this case study is the fact that users are recorded repeatedly, regardless of their expected enduring connection. A mobile phone session, for instance, may start by using the phone, and end when the user presses the lock button. Therefore, sessions always need to be aggregated into visits to the SL in question. This proposes an extra implication for SL4, since the visit may not be well presented by an aggregation.

Case Study 7: Architecture Room Q per Fifteen Minutes

A similar test has been executed with a temporal resolution of fifteen minutes. However, at the time of writing the data is not available in the database yet.

8.2. Data Analysis

Both the accuracy and representativeness of the processed data can be tested through validation tests. Therefore, data validation has been attempted in three different ways.

8.2.1. Data Validation

The final result of the filtering process is the amount of users per hour. To validate SL0 and SL1, all people entering and leaving a building can be counted. An overview of (semi) automatic counting methods can be found in Appendix IX. In order to simplify data validation, a building with a single entry is chosen. Furthermore, the TU Delft Library, as single entry facility, has itrack cameras installed by FMRE, to generate output on occupation.

Validation Test 1: itrack Camera's

Two *countcam* cameras are located at the ceiling above the building entrance (Figure 54). As the name implies, the *countcam* is a camera-based visitor counter that scans people moving across the camera view (itrack bv., 2015). Since this system is already installed, its output can be used as data validation. This reduces the time spent on less sophisticated ways of collecting ground truth, like manual counting. However, the reliability of the itrack system is unknown. Therefore, a validation test is performed on two different days through manual counting. The goal is to determine the reliability of the itrack cameras in counting the total persons entering and leaving the TU Delft Library. The full explanation of the test can be found in appendix X.

	11-May: 09.00-10.00	13 May: 9.00-10.00
Manual Main Entrance IN	283	204
Itrack Main Entrance IN	293	205
Deviation	*+3,5%*	*+0,5%*
Manual Main Entrance OUT	57	61
Itrack Main Entrance OUT	109	75
Deviation	*+91%*	*+23%*

Table 26: itrack data validation at the TU Delft Library.

As can been seen in table 26, the retrieved itrack data is considered viable for people entering the TU Delft Library, as a deviation below 5% is accepted. The data for people departing from the building is however deemed unreliable. A deviation of 23% could possibly be explained as validation errors, but a deviation of 91% cannot be justified. FMRE, as owner of the itrack scanners, is informed about this result.

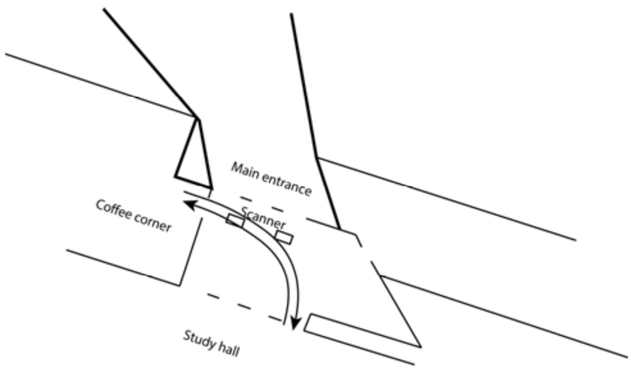


Figure 54: itrack Camera Locations at the TU Delft Library.

Validation Test 2: Manual Counting per Hour

The validation data retrieved to test the itrack cameras can also be used to validate the data output. However, as described in chapter 5, the minutes a user is inside a building are calculated instead of counting visits. For example, if a person is inside a building for twenty minutes, then three users visiting for twenty minutes is equal to a single working hour. Manual counting was performed over the course of two hours. For reliable data validation, the manual counting should be performed by the minute.

Validation Test 3: Manual Counting per Minute

To validate the results of the analysis process a manual counting session of two hours was performed at the library on 7-6-2016. The manual validation is performed through counting people who are entering or leaving the TU Library per minute. As the occupation describes the users per hour, counting entering and leaving persons per minute enables to aggregate minutes to incoming/outgoing users on an hourly basis. The results from the manual counting session can be seen in table 27.

	15.00-16.00	16.00-17.00
Manual counting (ground truth)	-24	-67
Processed data (database)	-36	-74
Deviation	+50%	+10,4%

Table 27: Overview of the manual counting at the TU Delft Library 07-06-2016.

The offset between the data from the database and the data from the manual validation can have different possible causes. Errors might be caused at the processed data due to aggressive filtering, not recording guest users and not recording users connected to the LAN network. The data from the manual validation can also contain errors; persons might be outside the TU Library, but their laptop or other devices are still inside and connected, persons or persons might be counted too early when they are not connected to the WiFi network yet.

Conclusion

The itrack cameras cannot be used to validate results. The manual counting is used as ground truth in the case study, which however is likely to contain errors. It should be concluded that this cannot be used as a proper ground truth. As a result, it cannot be stated whether the offset between the manual counting and processed data is to be blamed on either gross errors in manual counting, therefore approving the processed data, or errors in the captured or processed data.

Note: The effect of the offset between the manual validation and the database records on the complete database is dependent on the amount of people inside of the building. The manual validation was performed in the middle of the day by counting people leaving and entering; the results do not include the occupation of the building itself. Therefore, the results of this case study cannot be used to relate the offset to the occupation in the database.

8.2.2. Query Quality

Apart from the quality of the output itself, the queries used to generate said output can be optimized, both to produce reliable output and to run as efficiently as possible. In a general sense, the queries generate the desired output. However, when a query had to be updated based on quality control, only newly added data is processed accordingly. In order to improve already processed data, the *wifilog* needs to be processed in backup, which is then used to overwrite the final tables with the new data. As this process is not automatic, it should be improved.

Query 5

The section of the query used to remove the overlapping session durations can be improved. The current problem is that the duration of the current session is trimmed, if the start time of the next session is before the end time of the current session. In select cases, the session end of the next session also falls inside the session duration of the current session as shown in the right case in figure 55. This phenomenon leads to holes between session starts and shorter session duration.

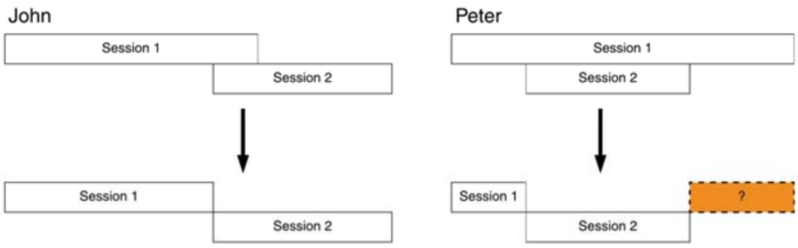


Figure 55: Two different cases in which session times are changed. The left case results in accurate trimming, while the right case results into missing session times

Query 6

The occupation in an hour together with the UFA determines the exploitation. However, when using the occupation to calculate the exploitation it is important to decide on how the occupancy is calculated as well. Should a user who is detected for only twenty minutes have the same weight as a user detected for the complete hour? The focus of calculating the exploitation is to see how much space is available per student in an hour, therefore simply counting the amount of visits in an hour per building is insufficient. Instead of counting visits, the minutes a visit overlaps with the hour are counted. For example, the occupation between 9:00 and 9:59 needs to be calculated, and there is a visit with a start time of 8:55 and a session duration of twenty minutes; the visit overlaps with the hour for those twenty minutes. For each hour, the overlap between hour and all visits is calculated. By dividing the total per hour by sixty, the amount of people is calculated.

Query 9

The data from *g01_floor_nuttigo* is also used to filter out floors from buildings not in scope. During the development of the query, it became clear that not all floor names adhere to a logical naming scheme. In some cases, room names were used instead of floor names. Other APs only had the default floor name, leading to the results being filtered out. Furthermore, many of the *maploc* entries contain spelling mistakes. A possible improvement would be to rename the locations of the APs in the system and for TU Delft system management to be more consistent when naming APs. Another issue is that some floors are unjustly renamed to a different floor name, because of a case in the query used to remove spelling mistakes. For example, there is the case that renames floor 1:

```
...when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%1%diep%'::text …
THEN '1e verdieping'::text
```

A floor named '9e-10e-11e verdieping' would also return true to the above case. This results into different double entries for certain floor names in the materialized view for buildings with grouped floors or more than nine floors. The issue was solved by adding an extra clause to the case but all data until 5-6-2016 was analysed using the old case.

8.3. Data Storage

The large amounts of dependencies make editing materialize views a job where multiple views have to be dropped before another can be edited. However, the ability to both rerun the query that created the view as well as the ability to have constant access to the resulting records without being forced to wait every time the view is called increases the performance of the database when data is being analysed (The PostgreSQL Global Development Group, 2016b). Further research could be done on testing the performance of the materialized views is the current setup against either function that creates tables for the analysis queries and drop them after completion, or against temporary tables.

When new data is added to the *wifilog*, the filtering and aggregating processes are automatically activated through a crontab, as described in paragraph 6.2. The new data will be added to three final tables: *g01_sl0_oeh*, *g01_sl1_oeh* and *g01_sl2_oeh*. The filtering and aggregating process takes up to thirty minutes for every day. In order to efficiently utilize computer power, the data storing and analysing processes are intertwined, so that visualizations can be created as quickly and easily as possible. Thus, the output is quickly accessible and the dashboard can run optimally. However, already processed data will not automatically be reprocessed by a crontab action or other abnormality. As described in paragraph 8.2.2, some levels were renamed in during the process, and therefore their connected data was not correctly processed. However, the adaptations made in the queries to account for this problem will only affect new data. Automating and fully implementing cleaning processes like this would be a subject for further research.

8.4. Communication

Recurring quality control on communication of results allows a streamlining of the project output to what the client expects. Thus, the client is enabled to adapt the set requirements to wisdom gained from preliminary results. An active back and forth in this quality control process allows for the production of both reliable and desired results.

8.4.1. Communication of Results

In communication with FMRE, both requirements for and feedback on produced outcome types and visualisations have been gathered (Valks, 2016c). A first requirement was the representation of exploitation in m^2/n. This output is generated on the database, as part of the analysis process. A distinction between use types was also requested, but it has been concluded that this information cannot be determined from the data. FRME voiced a preference for data on lower levels, though inquiry has adjusted this requirement to data with a high reliability. Therefore, the output on SL2 has been produced, however with the note that the data is not reliable. A requirement that arose at the end of the project was the availability of an interactive entry of ideal exploitation, as this value differs according to use type. Thus, the client can make more informed decisions, according to their knowledge of specific use of areas of interest.
Though the FMRE representative may be familiar with ideal exploitation values, not all recipients of the project output are expected to automatically read these numbers correctly. Therefore, the user friendliness of result communication has been increased by translating ideal values into percentages.

8.4.2. Visualisation of Data

A similar decision making process has been run with the MSc Geomatics mentor. This resulted in the incorporation of occupation in graphs, but not in 3D geometry, as occupation does not represent an actively comparable value. Another resulting decision was the incorporation of bar chart geometry, to allow for a user-friendly direct comparison of exploitation of all buildings. To increase the intuitive reading of this geometry, half the bar has been converted into an ideal value part.

8.5. Conclusion

From the discussed quality control subjects, the following can be concluded:

- With information currently available, a single device cannot precisely be located. An approximate horizontal location may be possible, but the floor cannot accurately be derived, as concluded from the executed case studies. Therefore, all occupation and exploitation numbers to be derived on SL2 will contain a very low reliability.
- The created dashboard provides knowledge on occupation, as number of people per hour per SL, and exploitation, as number of people per square meter per hour per SL.
- As tables with cleaned data are created within the database, the output is quickly accessible and the dashboard can run optimally. However, already processed data will not automatically be reprocessed by a crontab action or other abnormality.
- SL4 is not viable because it cannot be concluded from the data where or when exactly a change in occupation occurs.

9. Cross Cutting Topics

As mentioned in chapter 1, the conducted research should always take a few cross cutting topics into account. The implementation of these sub topics are woven into the project and can be found all through the report. This chapter specifically mentions some important references and provides additional information on each subject.

9.1. Privacy

Though the final product publishes aggregated numbers only, the privacy of all recorded users is an important aspect to take into account. The dashboard is developed on a localhost per machine, so despite its web-based development, accessibility of the output does not need to be privatized. However, processing of personal data, as done in the project, is a sensitive matter.

9.1.1. Personal Data

Personal data is defined as information about a natural person who can be identified through the dataset (European Union, 1995 Article 2). Since all *mac addresses* and *usernames* in the dataset are hashed, no natural person is directly identifiable in the database. However, once a single location and time for a certain person is known, this person's location can be easily tracked throughout the database, for both historical and future data. Therefore, the database handles personal data. An implication to the processing of personal data within this project is the lack of direct consent of data subjects. However, all data is gathered from Eduroam connections, which is only accessible to subjects registered at an educational facility. As the data is used for research within this educational facility and meant to improve the state of this educational facility, the processing of said personal data is allowed. However, at any given time a subject should be able to object processing of data related to him, given he has compelling legitimate grounds relating to his particular situation (European Union, 1995, Article 14)

9.1.2. Publication of Personal Data

Publication of information gathered from personal data not obtained from subjects directly should contain the following information (European Union, 1995, Article 11):

- The identity of the controller, as project in_sight and its team members.
- The purposes of the processing, as the grounds and goals of the project.
- The categories of data concerned, as records containing *username*, *mac address*, date, time and vicinity to a known Eduroam AP, where username and mac address are hashed and therefore not directly linkable to a natural person and the type of machine he / she is using.
- The recipients, as Geomatics, the project organizer, and FMRE, the project client.
- The existence of the right of access to and the right to rectify the data concerning one self.

The dashboard contains a few information pages, together mentioning all the required information regarding the publication of personal data.

9.2. Validity and Accuracy

The quality of the data can be assessed by measuring both the validity and the accuracy of the dataset. Here, validity is defined as well grounded and justifiable, and accuracy as conformant to a certain standard and free of error (Merriam-Webster, 2016)

9.2.1. Data Validity
The data validity is controlled by ensuring processing of all useful information. Data sources classifiable as threatening to this validity can be divided into two categories:

- Superfluous data. Many data sources have been made available, but not all are necessary to use. For example:
  - The data containing available floor space also covers technical areas, traffic areas and other rooms that do not directly influence the occupation of educational facilities. These need to be filtered out, before floor space calculations are made.
  - The TU Delft campus contains many buildings, not all considered in scope, as not all buildings are part of FMRE's portfolio. Furthermore, a project focus is on the educational facilities only, as these are certainly covered by the *wifilogs* and may provide a chance to distinguish staff and students in the database.
  - Not all dates covered by the *wifilogs* are of interest. As the project focusses on regular occupation rates, some special occasions may need to be filtered out. To which dates this concerns, may be retrieved from one of the concurrent synthesis projects.
- Deficit data. Some data necessary for the conducted research was not delivered. For example:
  - In order to drill down from SL2, a map of AP locations is necessary. Unfortunately, this information cannot be delivered. To retrieve this information surpasses the project scope.
  - The w*ifilog* only stores made connections between AP and device. If the records would store visibility rather than actual connection, users could be located more precisely and accurately, therefore enabling locating for SL2 and SL4.

9.2.2. Data Accuracy
Accuracy of data can be ensured by validating collected measurements according to ground truth. This was attempted through three validation tests, described in paragraph 8.2. Furthermore, the case studies in paragraph 8.1 prove that the range of a single AP spreads across multiple floors, diminishing the accuracy on SL2-4.

9.3. Representativeness

In part, representativeness is covered by data accuracy, as it should provide a comparison between registered data and ground truth. The representativeness however, takes a broader perspective, as it looks at parts of the data that could not be labelled with accuracy.

9.3.1. Underrepresentation
Parts of the loss in representativeness can be predicted, apart from technical accuracy. For instance, the *wifilog* only records known Eduroam users. Therefore, guests on campus will not appear in the data. Many special occasions will not be correctly recorded. Users also have the possibility to switch off their WiFi.

9.3.2. Overrepresentation
On the other hand, incorrect interpretation of data may cause overrepresentation of occupation. As mentioned in paragraph 5.3.3, a user may be recorded at different APs at the same time, due to the temporal resolution of the system. Furthermore, as the data is aggregated to a temporal resolution of an hour, the actual outcomes may be misleading, were they to be taken as harsh numbers (see paragraph 8.2). Though the project team has made an effort to make the output as representative as possible, the reliability of output figures should always be taken as a general approximation of the real world situation.

9.4. System of APs

The available system of APs has been evaluated from several sides. Paragraph 4.1 mentions the theoretical background of WiFi signals, and paragraph 4.2 the specific specifications of the system installed at the TU Delft. Together with the case studies in paragraphs 8.1 and 8.2, it can be concluded that the provided system of APs performs sufficiently for providing WiFi connection, though not for WiFi monitoring.

9.4.1. Sufficiency for SL0-1

For both SL0 and SL1, the system is evaluated to provide sufficiently reliable data, as generally speaking, the system captures a representative amount of users, and the distinction between facilities can be made rather accurately.

9.4.2. Sufficiency for SL2-4

As the case studies in paragraph 8.1 have proven, the derived data for SL2 is not sufficiently reliable, due to signal propagation and a certain denseness of APs. Had the APs only recorded users on the same floor, the system would be marked as sufficient, as the total representativeness of the data is considered satisfactory. With the system of APs marked as insufficient for SL2, a reliable drill down to SL4 cannot be made. A single AP covers an area much larger than a single room. Furthermore, most rooms are covered by several Aps. The distinction between different rooms should technically be possible, given certain types of information. However, the mentioned research and case studies, it has been concluded that no available information flow provides sufficient additional information to locate a user inside a certain room.

10. Evaluation

This chapter serves as an evaluation of the research subjects. The evaluation of occupation and exploitation will be discussed first. After this, related topics of the sub hypotheses are evaluated. The MoSCoW rules can be used for evaluation as well. The chapter finalises with answering the hypotheses.

10.1. Occupation

The occupation derived on SL0 and SL1 are considered both reliable and viable. The SL2 output is not considered reliable, though viable as a global indication of spread of people across campus. The values are aggregated per hour in the database, though a higher temporal resolution would lead to a higher accuracy of output values. However, the representativeness of values is considered sufficient for indicating occupation on SL0 and SL1. A further interesting indication would be the average values for multiple hours or multiple days, a computation that has not been included into the queries and therefore not retrievable from the dashboard.

10.2. Exploitation

As the exploitation is directly derived from occupation, the same limitations apply. The other source for exploitation is the UFA. As this information is provided by FMRE as main client, the data is considered reliable. Therefore the exploitation in m^2/n for SL0 and SL1 are considered reliable, and SL2 only viable as a global indication. The percentages are based on a rough average of ideal exploitation, derived from highly variable values. For example, the ideal exploitation for a lecture room is taken at 2 m^2/n and at 20 m^2/n for a laboratory. Therefore, the actual outcome in percentage is less reliable, though still providing a reliable exploitation ratio and comparison measure. This decrease in reliability is taken care of by adding an ideal exploitation parameter to the dashboard, where the client can use his knowledge of room types and their ideal exploitation to assess the depicted situation.

10.3. Reliability

The reliability of occupation and exploitation is not only dependent on direct influences, but also on the quality of the data capture, analysis and storage. The reliability of the data can be evaluated as suitable for SL0 and SL1. However, due to the AP system specifications, the data is not reliable for SL2-4. Recommendations have to be made for improvement of the system, so that WiFi monitoring can be used in a reliable manner for SL2-4 as well.

10.4. User Types

One of the goals of the project was to determine the occupation and exploitation of different user types, mainly as students and staff. As the usernames are hashed, no distinction is left in the type of username or the user type tag that belongs to it in the TU Delft database. Statistical information is available on the amount of students and staff subscribed. Though this information is considered reliable, it does not allow for a direct translation onto the data, as the active attendance of all subscribed users is not confirmed. As can be seen in table 28, the total recorded users comes rather close, but a difference of five hundred users, makes statistical derivations rather unreliable

.

Amount of Users Subscribed	Unique Recorded Netids
38111	37681

Table 28: Amount of Users from Statistical Information and the Data.

Another option for determining different user types would be to link users to a favourite AP, and check the use type for that area. Information on area use types has been made available by FMRE. However, as data derived at SL2-4 is not considered reliable, this method would be based on such gross assumptions, the output cannot be considered viable.

10.5. MoSCoW Rules

In the beginning of the project, a set of MoSCoW rules was defined for both prioritizing and evaluating the tasks to be done. This paragraph examines whether the stated goals are properly implemented.

Must	Should
• Contain sufficiently accurate information on the device connections to the APs to facilitate positioning on multiple SLs (SL). • Realise an online dashboard for easy access by FMRE to the processed data. • Have at least 1 SL covering the campus. • Have at least 2 SLs for analysing, filtering and visualising the data. • Give FMRE insight into problem areas on the campus by using filtering tools. • Have the implementation for historical data. • Contain research on the capabilities of the APs with the focus on positioning persons and room and floor boundaries.	• **Have a system that can be accessed by FMRE after the completion of the Synthesis project.** • Contain research on the range of the APs and the signal strength propagation. • Have analysis of chosen research and visualization tools.
Could	**Won't**
• Have partial implementation of SL3-4. • Have an application focussed on the student as a stakeholder. • Have campus wide coverage of SL0-2. • Have a 3D visualization of select SLs.	• Have campus wide implementation for SL3 and SL4. • Monitor the attendance of employees based on the exploitation of office space. • Provide a solution to the exploitation challenge for FMRE. • Have the implementation for real-time data.

Table 29: MoSCoW Rules Evaluation project.

10.5.1. Must / Should

Most of the goals set in both the Must and Should categories have been implemented, therefore indicating the advancement of the project. However, one of the main goals was not reached, as the positioning on SL2-4 has been evaluated to be unreliable. As this goal was marked as a Must, the derived data on SL2 was still published onto the dashboard as global indicator of the actual situation. At the moment of writing, the dashboard has not been made accessible to FMRE yet. However, this report will be available to them, as well as a user manual that has been for a one time set up. A restriction in accessibility is that a background program (Apache Tomcat) has to be run alongside the dashboard and the user should be connected to Eduroam with a TU Delft NetID, in order to gain access to the database.

10.5.2. Could / Won't

As can be seen in the rich picture (Figure 13), an option was to create an application focussed on the student as stakeholder. This application could be used by students in order to assess the occupation and exploitation of study locations and find a quiet place to work. As the database is only updated once every day, in an ideal situation, the dashboard cannot be implemented for real-time data. Furthermore, the queries for analysis take up some time. However, due to the crontab (Paragraph 6.2) and the active connection between database and dashboard, a semi-real time state could be of study locations and find a quiet place to work.

10.6. Results: Knowledge Gained

This section will shortly discuss some of the resulting values. Figure 56 shows two output tables for the data for two different weeks. The first table shows the data for week 3.10, which is an exam week. During the exam period, it is clearly visible in the results that the building with number 21, the TU Delft library, is overcrowded (shown in red). This corresponds to the general experience of students, as known by the members of project in_sight. During the exam period, the extended opening hours and quiet workspaces of the library attract many students. The fellowship (building number 66) is also quite crowded during the exam period.

At the beginning of a regular lecture week, it can be noted that the library remains very well used. The Aula Conference Centre (building number 20) and the Fellowship are even busier at the first week. The exploitation of the Faculty of ID is sometimes close to the optimal value of 14.5 m2/n. As noted earlier, the reliability of these values cannot be confirmed with certainty.

Day	Building no. / Time	03	05	08	12	20	21	22	23	30	31	32	34	35	36	37	38	43	45	60	61	62	64	66
Mo	09.00	6%	4%	73%	26%	8%	121%	17%	50%	43%	91%	36%	40%	47%	39%	50%	0%	1%	58%	69%	11%	48%	47%	40%
	14.00	3%	15%	40%	26%	52%	161%	28%	45%	47%	73%	54%	56%	56%	28%	39%	10%	1%	6%	21%	7%	50%	18%	66%
	19.00	1%	6%	16%	6%	22%	145%	14%	32%	10%	41%	11%	30%	25%	11%	30%	40%	1%	4%	4%	3%	8%	15%	85%
Tu	09.00	6%	14%	48%	17%	43%	97%	18%	30%	34%	34%	27%	41%	44%	22%	44%	5%	2%	3%	14%	4%	44%	3%	30%
	14.00	3%	25%	30%	32%	41%	161%	23%	40%	63%	50%	41%	58%	56%	30%	29%	9%	4%	5%	13%	12%	33%	48%	97%
	19.00	2%	6%	20%	5%	11%	161%	10%	23%	9%	32%	12%	31%	32%	11%	22%	30%	1%	5%	4%	2%	9%	9%	58%
We	09.00	2%	7%	58%	19%	9%	104%	11%	37%	18%	43%	56%	41%	32%	24%	40%	5%	1%	3%	5%	6%	38%	24%	54%
	14.00	2%	12%	43%	11%	25%	161%	19%	35%	27%	54%	50%	39%	73%	22%	40%	6%	1%	3%	54%	12%	35%	18%	97%
	19.00	0%	3%	15%	6%	21%	132%	10%	14%	5%	23%	12%	20%	54%	12%	25%	23%	1%	2%	4%	2%	8%	6%	39%
Th	09.00	12%	12%	32%	8%	15%	85%	13%	23%	23%	29%	25%	22%	39%	18%	35%	2%	2%	16%	11%	7%	30%	7%	13%
	14.00	2%	7%	30%	14%	26%	161%	29%	35%	43%	58%	40%	45%	50%	27%	30%	16%	3%	5%	24%	10%	44%	40%	58%
	19.00	0%	5%	16%	4%	12%	121%	9%	19%	8%	22%	10%	16%	11%	10%	19%	18%	1%	11%	11%	3%	7%	10%	25%
Fr	09.00	6%	4%	6%	4%	2%	26%	6%	10%	7%	13%	9%	14%	16%	8%	19%	2%	1%	1%	81%	3%	12%	14%	5%
	14.00	3%	2%	15%	4%	40%	26%	6%	13%	9%	12%	22%	12%	18%	11%	17%	9%	2%	4%	4%	3%	10%	10%	10%
	19.00	0%	2%	3%	2%	22%	22%	2%	3%	3%	4%	3%	3%	0%	2%	13%	5%	1%	1%	4%	2%	4%	3%	4%

Week 3.10 (11-15 April)

Day	Building no. / Time	03	05	08	12	20	21	22	23	30	31	32	34	35	36	37	38	43	45	60	61	62	64	66
Mo	09.00	0%	1%	11%	12%	52%	11%	5%	16%	8%	8%	13%	17%	4%	25%	2%	1%	1%	3%	8%	1%	16%	2%	181%
	14.00	0%	4%	14%	5%	50%	40%	9%	15%	12%	16%	27%	20%	14%	15%	2%	4%	2%	3%	9%	4%	30%	7%	97%
	19.00	1%	1%	3%	2%	3%	18%	2%	7%	1%	6%	6%	5%	1%	8%	9%	13%	1%	1%	4%	1%	6%	5%	21%
Tu	09.00	2%	16%	54%	36%	181%	8%	16%	60%	34%	54%	47%	41%	52%	36%	6%	18%	1%	5%	17%	5%	40%	23%	76%
	14.00	6%	20%	44%	38%	207%	121%	25%	40%	52%	63%	91%	66%	28%	60%	7%	8%	5%	11%	24%	10%	66%	32%	121%
	19.00	2%	5%	17%	6%	32%	81%	7%	20%	8%	17%	17%	17%	5%	17%	30%	28%	0%	8%	4%	3%	16%	6%	22%
We	09.00	6%	8%	29%	30%	91%	38%	13%	41%	33%	58%	104%	36%	47%	26%	6%	10%	1%	3%	14%	4%	38%	7%	112%
	14.00	5%	56%	32%	21%	104%	112%	22%	58%	48%	58%	73%	54%	52%	39%	9%	22%	1%	41%	13%	13%	43%	18%	104%
	19.00	7%	5%	12%	5%	9%	81%	5%	15%	7%	18%	16%	16%	4%	12%	32%	63%	1%	2%	12%	7%	11%	8%	27%
Th	09.00	2%	25%	41%	22%	132%	38%	16%	48%	28%	26%	41%	27%	40%	24%	7%	2%	1%	40%	18%	6%	34%	7%	207%
	14.00	3%	15%	29%	25%	121%	91%	26%	50%	54%	81%	69%	58%	47%	43%	9%	10%	2%	13%	7%	11%	44%	23%	242%
	19.00	1%	3%	22%	6%	91%	63%	10%	20%	7%	29%	13%	18%	3%	18%	30%	15%	1%	2%	4%	3%	12%	21%	50%
Fr	09.00	3%	13%	35%	19%	56%	34%	8%	25%	16%	47%	56%	35%	18%	33%	3%	8%	1%	1%	16%	1%	28%	5%	63%
	14.00	5%	9%	37%	44%	52%	91%	22%	40%	24%	36%	60%	43%	69%	35%	24%	26%	2%	9%	19%	5%	34%	35%	145%
	19.00	1%	4%	10%	10%	12%	37%	4%	7%	2%	17%	8%	10%	18%	9%	28%	32%	1%	3%	7%	2%	8%	7%	27%

Week 4.1 (18-22 April)

Figure 56: The exploitation numbers proportional to the ideal exploitation, an overview of two weeks. Red colors have an overexploitation, green is an ideal exploitation, blue is under exploitation.

10.7. Hypothesis Evaluation

This paragraph provides the evaluation of the hypotheses. As explained, there are three sub hypothesis defined that contribute to the overall hypothesis. The overall hypothesis is as follows:

H1: The alignment of occupation and exploitation of educational facilities **can** be indicated through WiFi monitoring on all SLs with adequate reliability.

H0: The alignment of occupation and exploitation of educational facilities **cannot** be indicated through WiFi monitoring on all SLs with adequate reliability.

The three sub hypotheses that contribute to answering the overall hypothesis are the following:

	H1	H0
1	The occupation **can** be determined with 80% or more reliability.	The occupation **cannot** be determined with 80% or more reliability.
2	The occupation and exploitation of use types 'educational' and 'staff' **can** be determined.	The occupation and exploitation of use types 'educational' and 'staff' **cannot** be determined.
3	The available system of APs **is** suitable for indicating occupation through WiFi monitoring.	The available system of APs **is not** suitable for indicating occupation through WiFi monitoring.

Table 30: Sub Hypotheses.

The design sprints, as explained in paragraph 3.4.3, are used to define proper implementations of the different SLs. After each design sprint, the three sub hypotheses are evaluated. In table 32 is an overview of the hypothesis evaluations shown.

	SL0 (Campus)	SL1 (Facility)	SL2 (Floor)	SL3 (AP)	SL4 (Room)
Sub-H 1	H1	H0	H0	H0	H0
Sub-H 2	H0	H0	H0	H0	H0
Sub-H 3	H1	H1	H0	H0	H0

Table 31: Matrix with the evaluation of the sub hypotheses.

Nearly all sub hypotheses are rejected on the five different SLs, thereby accepting the null-hypothesis (H0). The occupation is considered to be reliable on SL0, as at least 80% of the TU Delft students and staff is assumed to be connected to the Eduroam system. The occupation on all other SLs cannot be determined with a specified reliability, as is a result of the case studies explained in paragraph 8.1. The occupation and exploitation of the user types students of staff cannot be determined on any of the SLs, which is due to the lack of information about the amount of users and their location in the educational facilities. The AP system is evaluated to be suitable for the SL0 and Facility SL1), as the precise location of the APs is of less relevance in indicating occupation. The map of the locations of each AP was not available during this research, therefore the precise location could not be determined. Furthermore, the case studies explained in paragraph 8.1, illustrate the signal strength of single APs and conclude that the AP system is not suitable to specify the location of users on the floor or room level. An extensive explanation on the reasons behind the evaluation can be found in the sprint execution documents (Appendix V).

11. Conclusion

During the project, the requirements and specifications have been identified to enable the capturing of data through the use of WiFi monitoring. Project in_sight aimed to design and create a solution to process, analyse and store WiFi monitoring data. In addition, a smart and reusable workflow, which efficiently identifies the occupation and exploitation of the educational facilities at the TU Delft Campus using an automatically refreshed data structure, has been created. An online dashboard has been developed, to communicate the output on. Through this dashboard, the client is able to gain insight into privacy, validity, accuracy, reliability and the viability of the available AP system for the determination of occupation and exploitation at the TU Campus. This report provides a documentation of all the aspects necessary to achieve the goals to capture, analyse, store and visualise WiFi monitoring data. Furthermore, it includes the methods and concepts of the implementation as well all the technical details.

H1: The alignment of occupation and exploitation of educational facilities **can** be indicated through WiFi monitoring on all SLs with adequate reliability.

H0: The alignment of occupation and exploitation of educational facilities **cannot** be indicated through WiFi monitoring on all SLs with adequate reliability.

Reliability is considered as an important aspect during the research. Both the reliability of the data capture system, and the resulting exploitation and occupation were researched through seven test cases and three validation tests. The results from the case studies showed that the AP system is sufficient for the occupation and exploitation calculation of SL0 and SL1, and insufficient for calculation of occupation and exploitation numbers on SL2 through SL4.

The three sub hypotheses are rejected based on the performed research. It can therefore be concluded that the occupation cannot be determined with a 80% reliability for all SLs, the occupation and exploitation of use types 'educational' and 'staff' cannot be determined for all SLs, and the available system of APs is not suitable for indicating occupation through WiFi monitoring for all SLs. These sub hypotheses responses determine that the overall hypothesis is to be rejected. As a result, the research concludes that the alignment of occupation and exploitation of educational facilities cannot be indicated through WiFi monitoring on all SLs with adequate reliability.

With the obtained insight in the research subjects, the cross-cutting topics and after answering the hypothesis, the following research question can be answered:

To what extent can the alignment of occupation and exploitation of educational facilities on different scales be indicated through WiFi monitoring?

The research has presented a solution to determine the occupation and exploitation at the TU Delft on SL0 and SL1 using WiFi monitoring. The occupation and exploitation for SL2 is also taken into account, although the reliability is low and demands further research. Recommendations are given to improve the AP system for the use of WiFi localisation as well as topics encountered in the project which need further research, and possible improvements on the system created by team in_sight.

To conclude, The AP specifications of the Eduroam system are optimized for WiFi coverage, the performance of the system in the field of WiFi monitoring is significantly lower. Seven case studies were performed to test the reliability of the AP network, three validation tests were performed to test the accuracy of the occupation and exploitation results. An online dashboard with access to near historic data was created to provide an interactive visual interpretation of the occupation and exploitation of available space to the client, FMRE.

12. Recommendations

This chapter provides general remarks and recommendations on the research project and presents scenarios for future applications. The five scenarios may be used as focus for further research into occupation and exploitation through WiFi monitoring.

12.1. General Remarks for the Eduroam System

There are four general remarks to be made on the available Eduroam system, which are likely to increase the reliability of the provided outcomes.

- The naming of the APs in the *wifilog* should be more consistent. Currently, it contains entries like "Room K" and "2e Vedieping". Some of the inconsistencies are covered by using case detections in the database, though others still have to be checked manually.
- Visitors at TU Delft are not recorded by the *wifilog*, despite their influence on occupation and exploitation of the TU Delft campus. The problem could be solved by recording connection to the TU-Visitor network.
- A LAN user is not connected with an AP, and thus the connection is not recorded in the *wifilog*. Instead, the user is connected with a network switch, via a UTP wall socket. Possibly, the port of the network switch through which a user is connected can be recorded separately, and merged with the *wifilog*. A specific port number can then be mapped according to the location of the wall socket that is connected to the port at a remote location. This way insight into the exact location of the usage of LAN could be provided. This new table should contain at least a Boolean as to indicate whether or not a user is actually recorded in on the machine, so that devices not used will not count as users.
- One of the *usernames* is connected with a large amount of *macs*. This specific *username* represents anonymous@tudelft.nl, or "jfSJyL76FdN80YZhzvf5jhcob8VA6OqX/6Bp08PF+z4=". Currently, this account is filtered out as static device, as described in paragraph 5.3.1. In the month of April 2016, the account is recorded 223.161 times, representing 68.337 distinct *macs*. This is about 4.4% of the total records for that month. When performing a case to match the found *macs* in this anonymous account to existing *usernames* in the *wifilog*, multiple records can be redirected properly and used for analysis. These records correspond with 0,9% of the total records in the *wifilog*. In the current solution 3,5% of the entire *wifilog* is filtered out, as it is unknown whether the mac addresses describe human behaviour. This should be part of further research, as well as the effect of the that 0,9% is improperly discarded from further processing.

12.2. Recommendations for the in_sight System

There are six recommendations distinguished regarding the developed dashboard. The recommendations are the following:

- The queries used to analyse data can be improved in multiple ways: the queries used to identify users living on campus can be combined and the queries used to create visits should cut off on the date of the sessions. Also, cases exist where the removal of overlapping sessions is too aggressive in reducing session times. An example of this is shown in Figure 57.

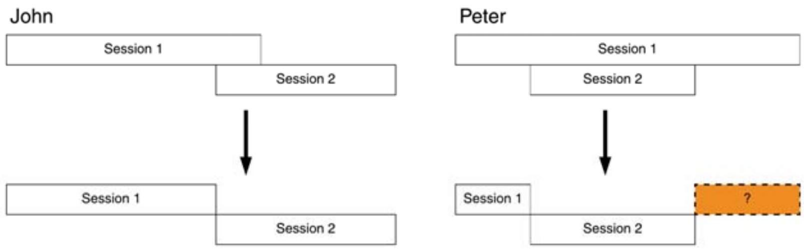


Figure 57: Example of Aggressive Session Time Reduction.

- If the database is prepared properly, the dashboard itself could be upgraded with the opportunity to view average values for multiple hours or days.
- With the current approach, it was not possible to determine a proper distinction between students and staff. Further research may include research into matching techniques based on not only occupancy, but also movement patterns as preliminary research of Pruitt and Grudin (2003) describes.
- The queries are prepared to automatically clean new incoming raw data. When however, multiple days are added at the same time, there is a possibility that cleaning tables will become too large. The queries could be adapted to process a permissible amount of data, perhaps thereby cleaning one day at the time.
- The scope of this research was to create a SDSS to be used by FMRE. An option could be to use the developed database system to create an application for student purposes, as described in paragraph 10.5.2.
- When queries for filtering and aggregating data are adapted, the processed days will not automatically be re-run with this new implementation. Therefore, only new data entered into the *wifilog* will be processed accordingly. Future research may include a solution to this.

12.3. Scenarios for Future Application

Currently, no implementation for SL4 has been made. This is due to the concluded low reliability of SL3 and the incapability of the provided system of APs to determine occupation on a more precise level. Further research may include additional methods to retrieve this information from the *wifilog,* and improving the reliability. Five scenarios are proposed that could be of interest to enhance the reliability on all SLs as well as enablement of the future implementation of SL4.

12.3.1. Scenario 1: Fixed AP Broadcasting Power

The current AP system transmits at a variable broadcasting power. The system of APs tries to reach an optimal coverage of the whole area, thereby varying the broadcasting power of the individual APs (Cisco, 2015). This has consequences for the reliability of the localisation of Eduroam users. Fixing the broadcasting power is therefore a scenario that can be considered (Figure 58).

Advantages

- A ratio can be established which describes the probability of users that are detected on the correct floor. As this ratio is based on the stored signal strength, the ratio can then be used to calculate the occupation per floor.
- The signal strength through walls, floors and other objects can be determined for individual APs varying over time based on the stored signal strength.

Disadvantages

- There is a high chance of signal overlaps when each AP is continuously broadcasting at full power. This effect can be decreased by determining the WiFi channels carefully as is also addressed by Lendino (2015).
- The net contribution of solely fixing the broadcasting power is considered relatively low.

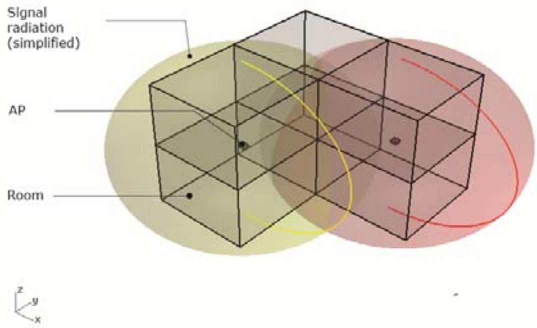


Figure 58: Fixed AP transmitting power: all APs transmit at their maximum power

12.3.2. Scenario 2: Save the variable broadcasting power of the individual AP

A second scenario would be to retain the variable broadcasting power setting and to include the broadcasting power of the individual AP in the database. When the broadcasting power is available, which is for example between 0-100%, a ratio can be determined that relates to the reliability of localising an Eduroam user. An example can be seen in Figure 59. When an Eduroam user is near the left AP (yellow), the reliability of being localised on the correctly is considered lower as when the Eduroam user is near the right AP (red). The ratio of the left AP (yellow) will therefore be low and the ratio of the right AP (red) rather high.

Advantages

- The users can still benefit from the current optimal WiFi coverage.
- No additional investments needed by FMRE.

Disadvantages

- Proper calculation of these ratios would have to be validated and should take error propagations from walls and floors into account.
- It is unknown if this broadcasting power is currently captured by the system, the integration of the broadcasting power could be problematic.

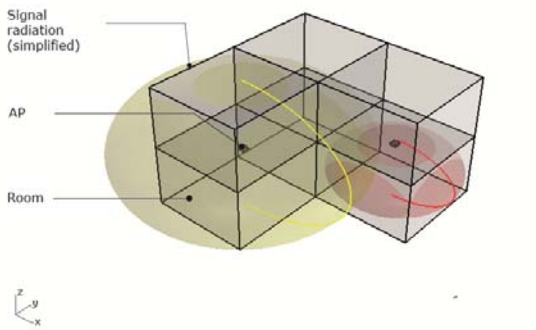


Figure 59: Variable AP transmitting power: APs transmit at a variable power level to give the best WiFi coverage possible.

12.3.3. Scenario 3: Store connection requests of detected APs to be used for trilateration

When searching for a connection the device of the Eduroam user, known as the client, and the AP communicate. The shared data that is shared is used to either accept or reject the device in connecting to the AP. This is a result of the load balancing between APs of the Cisco system (Cisco, 2011). A scenario option is to, right before connecting to the individual AP, capture and store the signal strengths of detected APs. As multiple signal strengths of different APs are recorded, trilateration possibilities increase.

Advantages

- Opportunity to increase accuracy of localization of Eduroam users.
- No additional investments needed by FMRE.

Disadvantages

- An accurate map with AP locations should be available, preferably visualized in 3D.
- A (complex) propagation model is to be created that takes the following aspects into account (Yang & Chen, 2009):
  - When using the signal strength, a proper implementation to handle the effected signal strengths due to walls, floors and other multipath obstacles is of importance in order to provide a reliable outcome.
  - The effects of the unknown variable broadcasting power on the reliability of the outcome.
- With trilateration the real position of a person can be calculated within 2 meters. This data is more private than the current data and raises extra privacy issues.

12.3.4. Scenario 4: Install additional APs with fixed broadcasting power in the specific room

A simple answer to a low reliability of the outcome, could be to consider the option of installing an AP in each room at the TU Campus (Figure 60). The broadcasting power could then be individually set to optimally cover the room, and this room only.

Advantages

- The signal coverage does not exceed the room and thus the Eduroam user that is connected to an AP is always located inside a room.

Disadvantages

- Before implementing this scenario, validation checks have to be performed in advance to ensure that this coverage can be realized optimal.
- This is an cost expensive option to be considered for FMRE, also when considering maintenance activities that have to be performed.
- Setting the proper broadcasting power has to be determined for each individual AP, which is a major technical complex task and will therefore be time expensive.

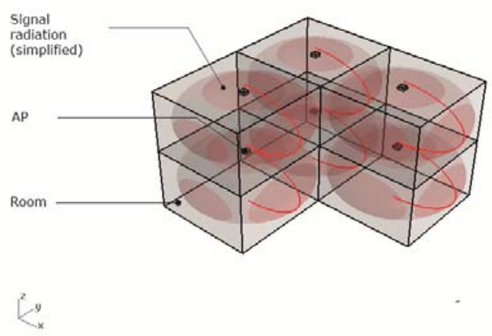


Figure 60: One AP per room with a limited broadcasting power would increase the accuracy up to room level.

12.3.5. Scenario 5: The use of additional equipment

The final scenario is the use of additional equipment.
Examples for extending the use of the current AP network with added WiFi monitoring technology are:

- Incorporate Angle of Arrival techniques in localisation. As described by Kotaru, Joshi, Bharadia, and Katti (2015) the angle of the WiFi-signal from the AP to the connected device can be used to enhance indoor localisation. As the current AP system is not suitable to capture this angle, additional equipment is needed. Examples are provided by companies as SpotFi and Arraytrack. A disadvantage of this technology, is that multipath has a large impact on the measurement (Cisco, 2014a).
- The addition of simple WiFi scanners. These scanners can be directed and will have a limited scanning range. The limited range can be used to locate users on SL4, though more WiFi scanners would be needed to cover the campus.
- As described by (Marcaletti, Rea, Giustiniano, Lenders, & Fakhreddine) it is possible to calculate the Time of Flight (ToF) of a WiFi signal (2014). The technique does not rely on the signal strength, but uses the speed of light as a distance measurement. The WiFi signal travels at a "speed that is close to the speed of light for most propagation media in typical indoor environments, and thus the signal propagation speed is fairly independent on the environment, obstacles, etc. "

Advantages

- Can be realized with small investments by FMRE.
- As the maintenance activities for the AP system involves renewals every 3-5 years (Valks, 2016b)the implementation can gradually be realised.

Disadvantages

- A large amount of pre-processing is needed to get a location, as is described by Marcaletti et al. (2014).
- The techniques described in the scenario have yet to be implemented in large scale projects, future research on their implementation is needed (Marcaletti et al., 2014, p. 18).

Examples for extending the use of the current AP by using other technologies to enhance outcomes are as follows:

- Installing counting cameras in areas of interest for FMRE.
- Installing heat cameras in areas of interest for FRME.

Advantages

- When there will be made use of other technologies there is less influence of the distortion of the WiFi signal.

Disadvantages

- Additional scanners will introduce the additional need for new infrastructure with associated costs and maintenance tasks.
- The additional equipment needs to be installed at the appropriate place for accurate results.

12.3.6. Final remarks

A combination of scenarios 5.1, angle of arrival technology, and 5.3, time of arrival, is the most optimal solution for indoor WiFi tracking on a high resolution while maintaining the functionality of the network. The application of both techniques takes places in the APs and can make use of the connection made between the AP and the device. By using both the angle of arrival technology and the time of arrival, only one AP is needed to locate a user as long as the location of the AP is known. During a connection request, the AP should request the device time needed to calculate the time difference.

13. References

Ackoff, R. L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis, 16*, p. 3-9.

antenna-theory.com. (2015). The Dipole Antenna. Retrieved from http://www.antenna-theory.com/antennas/dipole.php

Binnekamp, R. (2014, 18-04) *Aligning Real Estate Demand & Supply*. Manegerial Problem Solving in the Built Environment, TU Delft.

Cisco. (2011). Aggressive Load Balancing on Wireless LAN Controllers (WLCs) Release 6.0.188.0 and Later Configuration Example. Retrieved from http://www.cisco.com/c/en/us/support/docs/wireless/4400-series-wireless-lan-controllers/113160-aggressive-load-balancing-clients-00.html

Cisco. (2013). Cisco Aironet 1250 Series Access Point Hardware Installation Guide. OL-8247-03. Retrieved from http://www.cisco.com/c/en/us/td/docs/wireless/access_point/1250/installation/guide/1250HIG_TD-Book-Wrapper/125h_c1.html

Cisco. (2014a). Chapter: Location Tracking Approaches. Retrieved 13-06-2016 http://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Mobility/WiFiLBS-DG/wifich2.html

Cisco. (2014b, 22-1-2008). Cisco Aironet 1250 Series Access Point Radio Upgrade Instructions. Retrieved from http://www.cisco.com/c/en/us/td/docs/wireless/access_point/1250/radio/upgrade/guide/1250rdup.html

Cisco. (2014c). Cisco Aironet 1600/2600/3600 Series Access Point Deployment Guide. Retrieved from http://www.cisco.com/c/en/us/td/docs/wireless/technology/apdeploy/Cisco_Aironet.html#pgfId-44908

Cisco. (2014d, 20-11-2014). Cisco Aironet Antennas and Accessories Reference Guide. Retrieved from http://www.cisco.com/c/en/us/products/collateral/wireless/aironet-antennas-accessories/product_data_sheet09186a008008883b.html

Cisco. (2015). Wireless LAN Design Guide for High Density Client Environments in Higher Education. Retrieved from http://www.cisco.com/c/en/us/products/collateral/wireless/aironet-1250-series/design_guide_c07-693245.html

Cisco. (2016). Cisco Wireless LAN Solutions [Press release]. Retrieved from http://www.cisco.com/c/dam/en/us/products/collateral/wireless/aironet-1140-series/C45-614928-00_Channel_AAG_Cisco_Unified_Wireless_Network_Portfolio.pdf

Commotion Wireless. Learn Wireless Basics. Retrieved from https://commotionwireless.net/docs/cck/networking/learn-wireless-basics/

den Heijer, A. (2011). *Managin the university campus; Information to support real estate decisions*. Dleft: Eburon.

European Union. (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Union L281* (pp. 31-50). Brussels: European Union.

G. Bellinger, D. Castro, & Mills, A. (2004). Data, Information, Knowledge, Wisdom. Retrieved from http://www.systems-thinking.org/dikw/dikw.html

Google. (2015). *Design Sprint Methods , Playbook for start ups and designers*. Retrieved from https://developers.google.com/design-sprint/downloads/DesignSprintMethods.pdf

Hillson, D., & Simon, P. (2007). *Practical Project Risk Management: The ATOM Methodology*: Management Concepts.

IIBA. (2016). *A Guide to the Business Analysis Body of Knowledge* Toronto, Canada: International Institute of Business Analysis.

itrack bv. (2015). In surfcountcam.jpg (Ed.): itrack bv.
Kotaru, M., Joshi, K., Bharadia, D., & Katti, S. (2015). *SpotFi: Decimeter Level Localization Using WiFi*. Paper presented at the Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, London, United Kingdom.
Lemmens, M. J. P. M. (1991). GIS: the data problem.
Lemmens, M. J. P. M. (2011). *Geo-information;Technologies, Applications and the Environment* (5 ed.). Dordrecht: Springer.
Lendino, J. (2015). How to boost your Wi-Fi speed by choosing the right channel. Retrieved from http://www.extremetech.com/computing/179344-how-to-boost-your-wifi-speed-by-choosing-the-right-channel
Marcaletti, A., Rea, M., Giustiniano, D., Lenders, V., & Fakhreddine, A. (2014). *Filtering Noisy 802.11 Time-of-Flight Ranging Measurements*. Paper presented at the The 10th ACM International Conference on emerging Networking EXperiments and Technologies, Sydney, Australia. http://conferences2.sigcomm.org/co-next/2014/CoNEXT_papers/p13.pdf
Mautz, R. (2012). *Indoor Positioning Technologies*: Institut für Geodäsie und Photogrammetrie.
Merriam-Webster. (Ed.) (2016) (Online ed.). Merriam-Webster
Pruitt, J., & Grudin, J. (2003). Personas: Practice and Theory (pp. 15).
Statista. (2014). Average number of connected devices used per person in selected countries in 2014 (Statistic). Retrieved 14-06-2016 http://www.statista.com/statistics/333861/connected-devices-per-person-in-selected-countries/
Technische Universiteit Delft. (2010). TU Delft Campus map. In http://nsweb.tn.tudelft.nl/uploads/Plattegrond_2010.pdf (Ed.): Technische Universiteit Delft.
The Open Group. (2013). Base Specifications Issue 7. Retrieved 10-6-2016, from The Open Group http://pubs.opengroup.org/onlinepubs/9699919799/utilities/crontab.html
The PostgreSQL Global Development Group. (2016a). PostgreSQL 9.5.3 Documentation. Retrieved 10-6-2016, from The PostgreSQL Global Development Group https://www.postgresql.org/docs/9.5/static/sql-refreshmaterializedview.html
The PostgreSQL Global Development Group. (2016b). PostgreSQL Documentation, 38.3. Materialized Views. Retrieved 16-05-2016, from The PostgreSQL Global Development Group http://www.postgresql.org/docs/9.5/static/rules-materializedviews.html
Valks, B. (2016a, 18-04-2016). *Campus Development TU Delft - Facilities Management & Real Estate*, Technische Universiteit Delft.
Valks, B. (2016b, 13-06). [Feedback on Technical Presentation].
Valks, B. (2016c, 25-04-2016). [Personal Communication].
van der Spek, S., & Verbee, E. (2016). *Monitoring Flows and Occupation Patterns in Wi-Fi* (1.0 ed.). Technische Universiteit Delft: Technische Universiteit Delft.
Yang, J., & Chen, Y. (2009, Nov. 30 2009-Dec. 4 2009). *Indoor Localization Using Improved RSS-Based Lateration Methods*. Paper presented at the Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE.

Appendices

APPENDIX I: STAKEHOLDERS III

APPENDIX II: RESEARCH SCOPE IV

APPENDIX III: RISK REGISTER V

I.I. RISK IDENTIFICATION V
I.II. RISK ANALYSIS: DETERMINE RISK INTERFACES V
I.III. RISK RESPONSE PLANNING V
I.IV. RISK REGISTER VI

PROJECT MANAGEMENT EVALUATION VIII

I.V. HOW DID YOU DO THE PROJECT MANAGEMENT IN YOUR TEAM? VIII
I.VI. HOW DID YOU DEAL WITH RISKS? X
I.VII. DID YOU CHANGE THE APPROACH DURING THE PROJECT? DID YOU EXPERIMENT? IF NOT, WHY NOT? IF SO, HOW? X
I.VIII. HOW DID YOU DIVIDE THE DIFFERENT ROLES IN YOUR TEAM? XII
I.IX. TEAM EFFORT XIII
I.X. HOW CAN GROUPS IN THE NEXT SYNTHESIS PROJECT BENEFIT FROM YOUR EXPERIENCES? WHAT WOULD YOU HAVE DONE OTHERWISE IF YOU COULD START AGAIN? XIII

APPENDIX IV: SPRINTS XIV

I.XI. SPRINT CONCEPT XIV
I.XII. SPRINT 1: CAMPUS / FACULTY XV
I.XIII. SPRINT 2: FLOOR / AP XVIII
I.XIV. REPORT XXI
I.XV. SPRINT 3: FLOOR / DASHBOARD XXI
I.XVI. SPRINT 5: ROOM / DASHBOARD XXVI
I.XVII. REPORT XXIX

APPENDIX V: WORK BREAKDOWN STRUCTURE (WBS) XXX

APPENDIX VI: FILTERING XXXI

I.XVIII. QUERY 1: IDENTIFYING STATIC DEVICE USERS XXXI
I.XIX. QUERY 2-4: IDENTIFYING USERS LIVING ON CAMPUS XXXI
I.XX. QUERY 5: VISIT CREATION AND FILTERING OF USERS LIVING ON CAMPUS FOR SL1 XXXII
I.XXI. QUERY 6: CALCULATING OCCUPATION AND EXPLOITATION FOR SL1 XXXIII
I.XXII. QUERY 7: CALCULATING OCCUPATION AND EXPLOITATION FOR SL0 XXXIV
I.XXIII. QUERY 8: VISIT CREATION AND FILTERING OF USERS LIVING ON CAMPUS FOR SL2XXXIV
I.XXIV. QUERY 9: CALCULATING OCCUPATION AND EXPLOITATION FOR SL2 XXXVI

APPENDIX VII: KEY PERFORMANCE INDICATOR (KPI) FRAMEWORK XXXVIII

I.XXV. ACTIVITY INDICATORS XXXIX
I.XXVI. PRODUCT INDICATORS XL
I.XXVII. RESULTS XLI

APPENDIX VIII: VALIDATION TECHNIQUES XLIV

I.XXVIII. SYSTEMS USING CONTACT-TYPE COUNTERS XLIV
I.XXIX. SYSTEMS USING SENSORS XLIV
I.XXX. VISION-BASED SYSTEMS USING CAMERAS XLV
I.XXXI. MANUAL COUNTING XLV
I.XXXII. CONCLUSION XLV

APPENDIX IX: ITRACK VALIDATION XLVI

I.XXXIII. ASSESSMENT MODEL XLVI
I.XXXIV. COLLECT AND ANALYSE DATA... XLVII
I.XXXV. EVALUATE RESULTS... XLVII
I.XXXVI. CONCLUSION XLVII

APPENDIX X: LOGGING LAN USERS.. XLVIII

REFERENCES XLIX

Appendix I: Stakeholders

The involved stakeholders can be seen in Figure I.

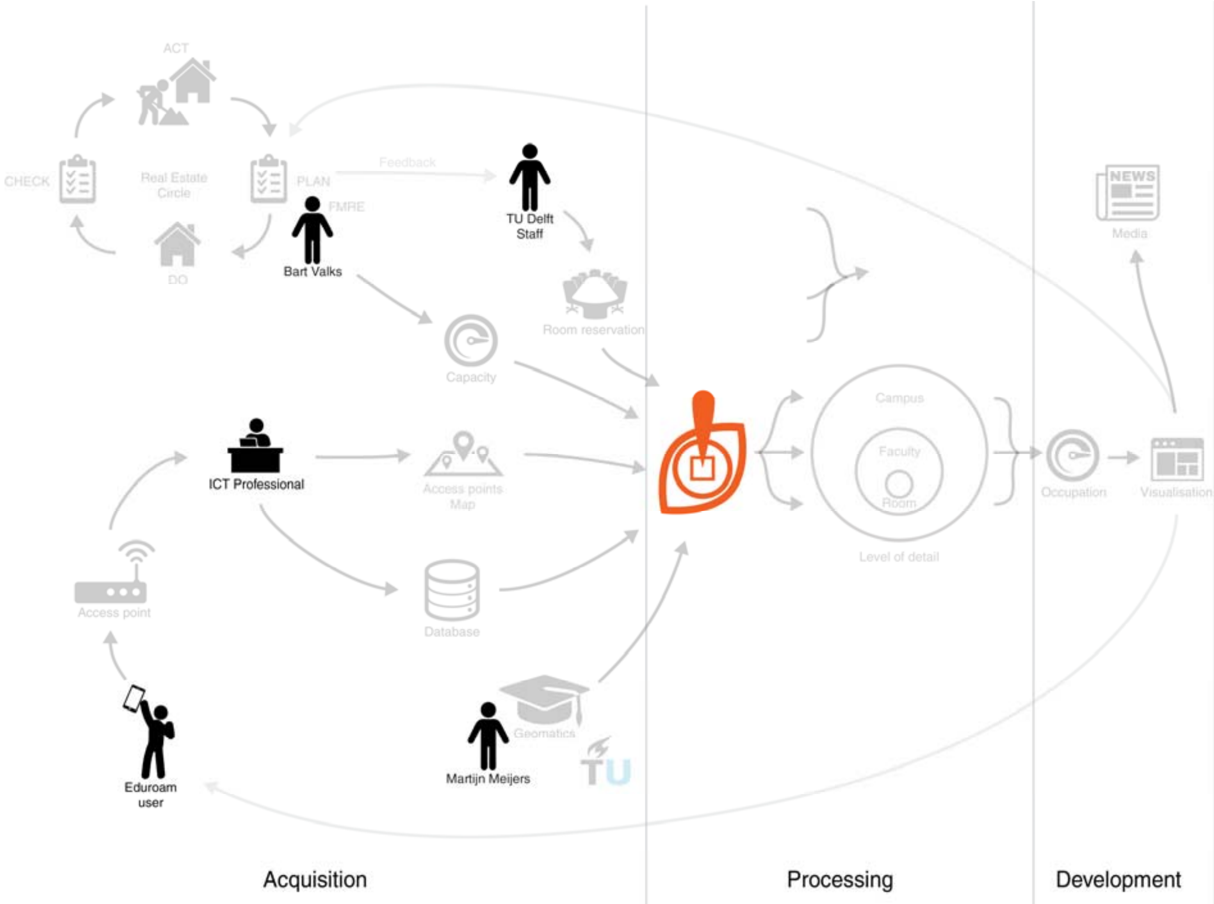


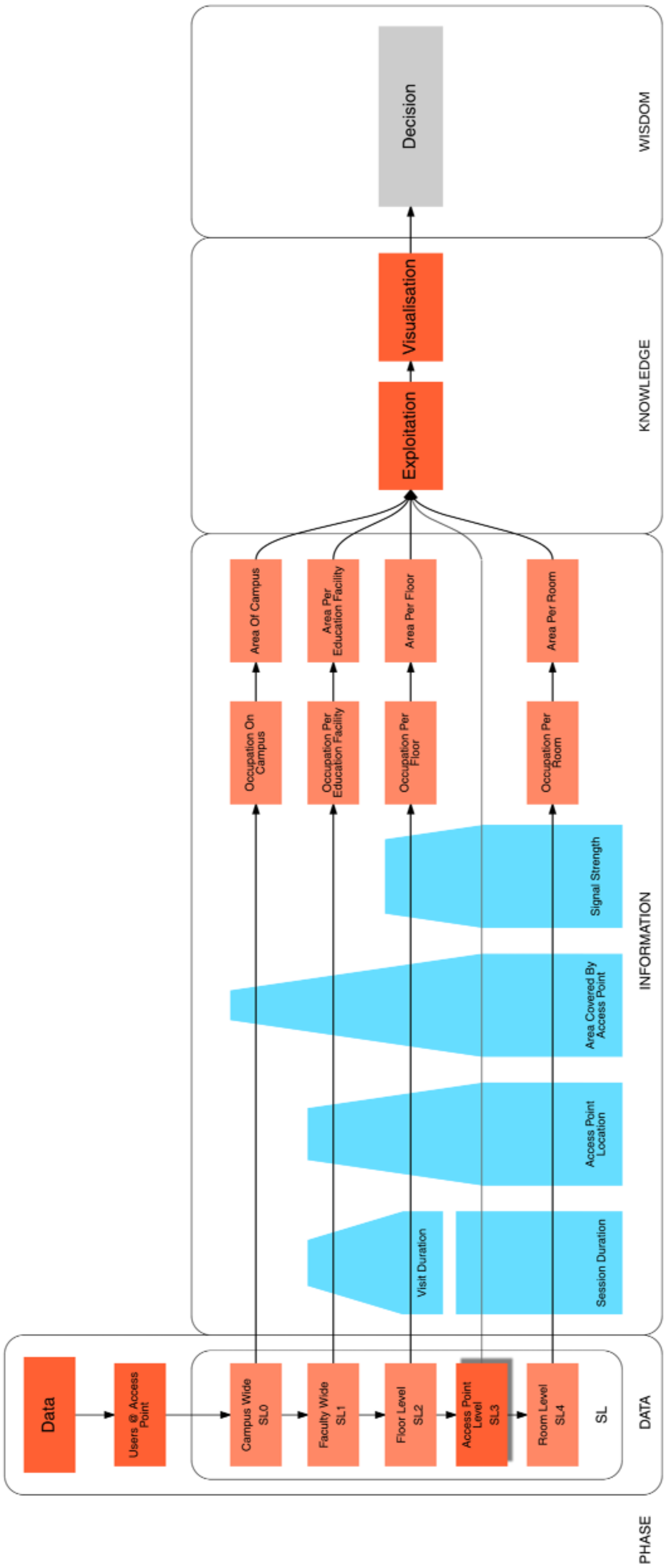
Figure I: Stakeholders Highlighted in Rich Picture.

The main stakeholders and their roles in the project are as follows:

- FMRE, main client:
  - Bart Valks, representative as main client.
- TU Delft Geomatics Department:
  - Martijn Meijers, team mentor.
  - Stefan van der Spek, course coordinator.
  - Edward Verbree, course coordinator.
- Available experts:
  - Wilko Quak, database expert. (ICT professional)
  - Lorenzo Dalla Corte, data protection expert.
  - Robert Voûte, project management expert.

An additional stakeholder is the TU Delft ICT department. An ICT professional provides the data in a database through a secured connection. He determines what data is available for the project and under which conditions. Two other stakeholders are the students and staff of the TU Delft. The stakeholders are presumed to be out of the research scope as these groups should only indirectly benefit from the decisions of FMRE. They could however directly benefit from separated applications specified to their needs as users of the researched spaces.

Appendix II: Research Scope



Appendix III: Risk Register

A risk can be defined as the probability of an adverse outcome and the severity of the consequences if the outcome does occur. In this chapter, the aim is to elaborate on the important risks in the research project. Focus is on the three steps in performing risk management: Identify, analyse and respond to risks (Figure II). A first step is to identify possible risks. Subsequently, the probability and impact of the risk has to be analysed. The final step is to define a proper risk response to handle the risk. Examples are described by the ATOM method, which describe to accept, reduce, avoid or transfer risks (Hillson & Simon, 2007)

Figure II: Proper Risk Management Consists of the Three Steps to Identify, Analyse and Respond to Risks.

I.I. Risk Identification

The objective of risk identification is the early and continuous identification of events that have negative impacts on the project, if they appear. These impacts can be on performance, outcomes or goals defined. In this research project, the risk identification is focused on the following aspects:

- Technical assessment, as system validity and reliability through representativeness of users, APs and accuracy of data.
- Organizational backgrounds.
- Privacy, as the sensitivity towards personal data.
- External grounds, mainly by the involved stakeholders.

I.II. Risk Analysis: Determine Risk Interfaces

The existence of multiple interfaces increases risk during the project. Interfaces can be described as the aspects in a project where independent subsystems connect. Four types of interfaces can be distinguished (Figure III). In this project, the assessment of the technical risks is mainly interesting (van der Spek & Verbee, 2016).The discovered interfaces call potential risks and predefine a complete risk register .

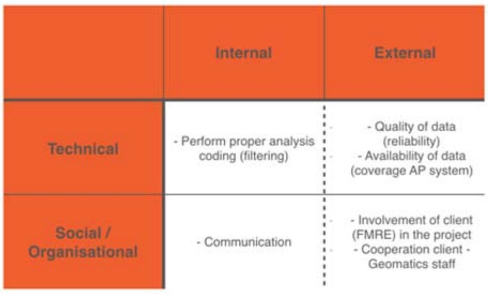


Figure III: The Identification of Interfaces in the Interface Matrix as based on Leijten (2014).

I.III. Risk Response Planning

In response to the identified risks, options and actions are defined, to enhance opportunities and reduce threats. Four response strategies can be distinguished (PMI Standards Committee & Project Management Institute, 1996).The impact of a risk on the project objectives and goals determine the mitigation response (Figure IV).

I.III.I. Avoid

By avoiding a risk, the project plan is adjusted to eliminate the risk in order to protect the project objectives from its impact. Some risk events arise early enough in the project to be avoided by clarifying requirements, obtaining information, improving communication, or acquiring expertise.

I.III.II. Transfer

Risk transfer is finding a shift in the consequence of a risk towards a third party, and thereby ownership of the response. Transferring does not eliminate a risk, but moves the responsibility for its management to another party.

I.III.III. Reduce

When reducing a risk, the aim is to reduce the probability or consequences of an adverse risk event to an acceptable threshold.

I.III.IV. Accept

By accepting certain risks, the project team decides not to change the project plan to deal with a risk or is unable to identify any other suitable response strategy.

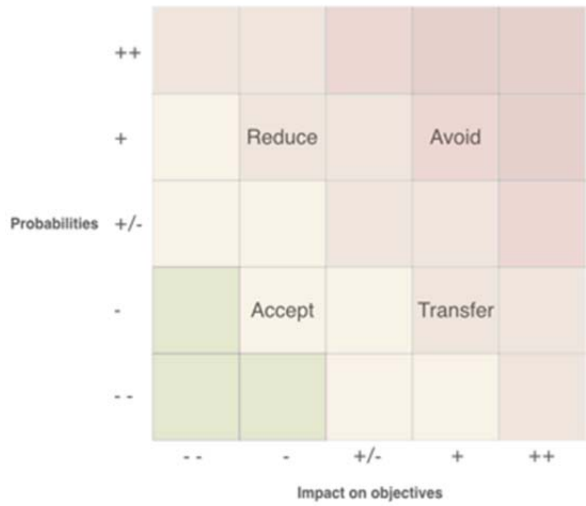


Figure IV: The risk response matrix is used to either reduce, avoid, accept or transfer risks.

I.IV. Risk Register

"Risk management is the identification, assessment and prioritization of risks followed by coordinated and economical application to minimize, monitor and control the probability and impact of unfortunate events or to maximize the realization of opportunities" (Hubbard, 2009, p. 49)

A risk register is drawn, including seventeen risks. Its legend is explained in Table 3. It should be noted that identifying, assessing and mitigating all possible risks of the project is almost impossible. However, the existence of many risks should be acknowledged and not be hidden, regardless of the fact that the project could become unattractive to execute (Vrancken, 2014).

Aspect	Sub-aspect	Defined by
Category		Technical, Organizational, Political, Financial, External, Legal
Risk Description	Cause	A fact
	Risk event	Something that might happen
	Consequence	Impact on project promise
	Project promise	Scope, Cost, Time, Quality
Pre Response Assessment	Probability	Very low, low, moderate, high, very high
	Impact	Very low, low, moderate, high, very high
Risk Response	ATOM	Avoid, Transfer, Reduce, Accept, Exploit, Share, Enhance
After response Assessment	Probability	Very low, low, moderate, high, very high
	Impact	Very low, low, moderate, high, very high

Table I: Definition of aspects in risk register.

	Category	Risk description				Pre-response Assessment		Risk response		Post-response Assessment	
Risk		Cause	Risk event	Consequence	Project promise	Probability	Impact	Description	ATOM	Probability	Impact
1	Political	- Sudden conflict of interest between the FMRE and Geomatics staff	Unclear definition of project goal and/or deliverables	Ambiguity about scope for project insight	Scope	LOW	VERY HIGH	Weekly feedback moment mentor and FMRE client	Reduce	VERY LOW	HIGH
2	Political	- Shift in client demand	FMRE withdraws as main stakeholder for the project	Loss of information flows	Quality	LOW	HIGH	Involve FMRE actively and emphasize the added value of their output providance and project insight's project output	Reduce	VERY LOW	HIGH
3	Legal / Financial	- Raise of privacy issues to perform data validation by facial recognition - Expensive to place validation equipment	Automatic data validation (of equipment) not possible	Manual validation of data needed	Quality, Time	MODERATE	HIGH	Search for semi-automatic and manual data validation options	Reduce	MODERATE	MODERATE
4	Legal	- Raise of privacy issues	Protests (of either students/staff) to use their eduroam data	Eliminate personal data from database Representativeness of dataset affected	Quality, Time	LOW	LOW	As the total amount of users of the eduroam network is on average 18.000 users a day, the few eduroam users that are aware will not affect the output quality	Accept	LOW	LOW
5	Legal	- Raise of privacy issues	Endresult may not be published due to private legislation	FMRE cannot obtain knowledge from project insight	Scope	MODERATE	MODERATE	The progress of project insight will not be affected by the risk as the impact is mainly for the client	Transfer	LOW	MODERATE
6	Financial	- Desire to present output in an online dashboard (3D)	- High costs of a desired operations dashboard (client/server)	There is no online 3D operations dashboard for this projects' dynamic data	Scope	MODERATE	MODERATE	When no open source 3D supported operations dashboard is available, a subsidy can be optional.	Accept	MODERATE	MODERATE
7	Financial	- Need to validate data gathered from Itrack system (TU Library) and AP system	Sufficient data validation requires budget for additional equipment	No sufficient data validation can be performed	Quality	MODERATE	MODERATE	The potential loss of quality will be appointed to the stakeholders	Transfer	LOW	MODERATE
8	Organizational	- Illness of mentor/FMRE	Absence of mentor/FMRE	Insufficient feedback during the project	Quality	MODERATE	HIGH	The (un)availability of stakeholders cannot be changed, though contact via email can be stimulated	Avoid	MODERATE	HIGH
9	Organizational	- Illness of team member - Quitting of team member	Absence of one or multiple team members	Insufficient participation of team member, possible team conflicts, increase of effort other team members	Time, Scope, Quality	LOW	VERY HIGH	Team building events and evaluations for the team and personal level	Reduce	VERY LOW	VERY HIGH
10	Organizational	- Critical path of project planning causes delays	Delay in overall planning	Out of time deliverables, possible quality issues and cutting corners	Time, Quality	MODERATE	MODERATE	Weekly check with the Gantt-chart	Reduce	LOW	MODERATE
11	Organizational	Capabilities of project in_sight are not sufficient to create a certain dashboard	There is no online 3D operations dashboard with dynamic data	Only 2D maps representation of analysed data	Scope, Quality	MODERATE	MODERATE	The only response that can be taken is to actively seek solutions within our knowledge area	Accept	MODERATE	MODERATE
12	Organizational	- Unclear communication between group members - Unequal distribution in work load	Personal conflicts	Productivity of team members (and thus project) decreases	Quality, Time, Scope	LOW	VERY HIGH	Weekly team events such as lunch, dinner, coffee time :)	Reduce	VERY LOW	VERY HIGH
13	Organizational	- Team member deprives the opportunity for other team members to perform - Team member unwilling to perform	Unequal distribution in work load throughout the project	Productivity of team members (and thus project) decreases	Quality, Time, Scope	MODERATE	HIGH	Personal contact within team to stimulate team members to equally participate during the project	Avoid	LOW	MODERATE
14	Technical	- Automatic / manual validation of provided data not possible within time span of project	No accurate conclusive information about the quality of the data output can be provided	Reliability of data output cannot be guaranteed to client	Quality	MODERATE	VERY HIGH	Manual validation at an early stage, foresee client with issues regarding this risk and involve client in validation process	Reduce	LOW	VERY HIGH
15	Technical	- The map of AP locations cannot be provided by the ICT specialist	The access point (AP) locations are not available on a map	Mandatory decrease in scope due to manual validitation of AP locations	Time, Scope	HIGH	VERY HIGH	Manual validation at an early stage, foresee client with issues regarding this risk and involve client in validation process	Reduce	LOW	VERY HIGH
16	Technical	- Technical issues with the online data server	Loss of data, broken connection with data server	No access to database	Quality, Time, Scope	LOW	VERY HIGH	Weekly database export to local computer	Reduce	VERY LOW	VERY HIGH
17	Technical	- Right to object of students/staff to opt out to the eduroam system	Manually eliminate data entries of students/staff	Decrease of data quality	Quality, Time	LOW	LOW	The few eduroam users that are aware of this project and that do object can be taken out of the database	Accept	MODERATE	LOW
18	Technical	Filtering of data is not valid or justified	Create improper conclusions	The quality of the output of the project is to be challenged	Quality	HIGH	HIGH	SQL check by other team members, perform filters step by step	Reduce	LOW	HIGH

Table II: Risk register identifies 18 risks, which are captured in the categories political, legal, organisational and technical.

Project Management Evaluation

How did you do the project management in your team?

Within the first phase of the project, the necessary roles within the team have been defined. As the project deals with two main stakeholders, thus dividing requirements and constraints to the project into two subdivisions, the team has been divided into two partitions as well (Figure V). The specific team roles and responsibilities are explained at question 4. After defining the team roles, the individual team members were assigned. As the project team consists of five people, two members take part in the development team, aided by the rest of the team when necessary.

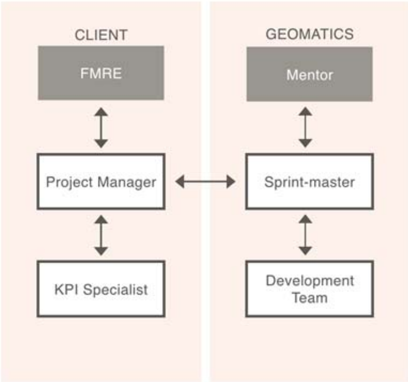


Figure V: Organisational Breakdown Structure (OBS) for Project in_sight.

The process itself is divided into design sprints (based on the Google approach). Each sprint is designed to explore all options available, and then choose the optimal way to continue (Figure VI). As the research scope divides the process into several Spatial Levels (SLs), each sprint contains a focus on one or two of these SLs (Table III). The SLs are then evaluated according to the hypotheses and vice versa. An example of this sprint planning is depicted in figure VII. The total planning of sprints, including delay, is depicted in table IV.

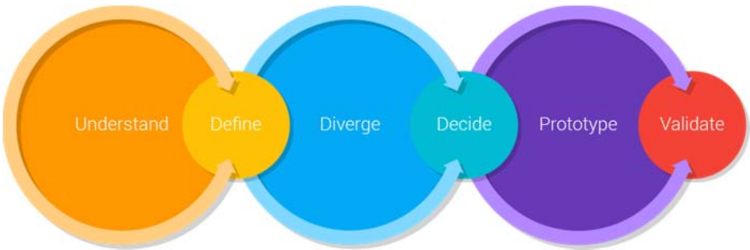


Figure VI: The six steps of a Design Sprint: Understand, define, diverge, decide, prototype and validate.

	SL0 (Campus)	SL1 (Faculty)	SL2 (Floor)	SL3 (AP)	SL4 (Room)	Evaluation hypothesis
Hypothesis	Sprint 1	Sprint 1	Sprint 2/4	Sprint 2	Sprint 3/5	
Sub-H 1	Sprint 1	Sprint 1	Sprint 2/4	Sprint 2	Sprint 3/5	
Sub-H 2	Sprint 1	Sprint 1	Sprint 2/4	Sprint 2	Sprint 3/5	
Sub-H 3	Sprint 1	Sprint 1	Sprint 2/4	Sprint 2	Sprint 3/5	
Evaluation of SL						

Table III: Division of SLs and Hypothesis Assessment for the Design Sprints of Phases 2 and 3.

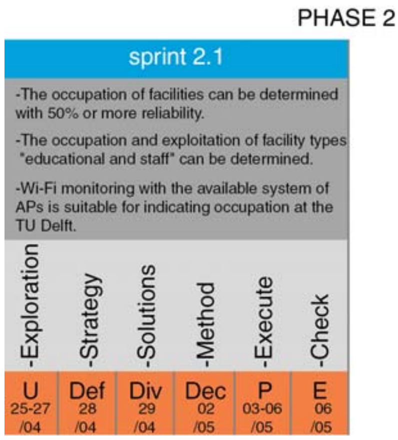


Figure VII: Example of design sprint 2.

Week	1	2	3	4	5	6	7	8	9
	Phase 1	Phase 2				Phase 3			Phase 4
	Project Plan								
		Sprint 1: Campus / Faculty							
				Sprint 2: Floor					
					Report				
						Sprint 3: Floor			
							Sprint 4: Room / AP		
								Report	
									Sympo-sium

Table IV: Sprint Planning and Progress. Orange Line: Deadline. Pink Block: Delay.

The practical planning of a sprint is first of all based on the Work Breakdown Structure (WBS), defining global tasks necessary for a complete research process (VIII). The tasks are then discussed, further defined and assigned. The practical output of this process can be found in the Final Technical Report (Appendix V: Sprints).

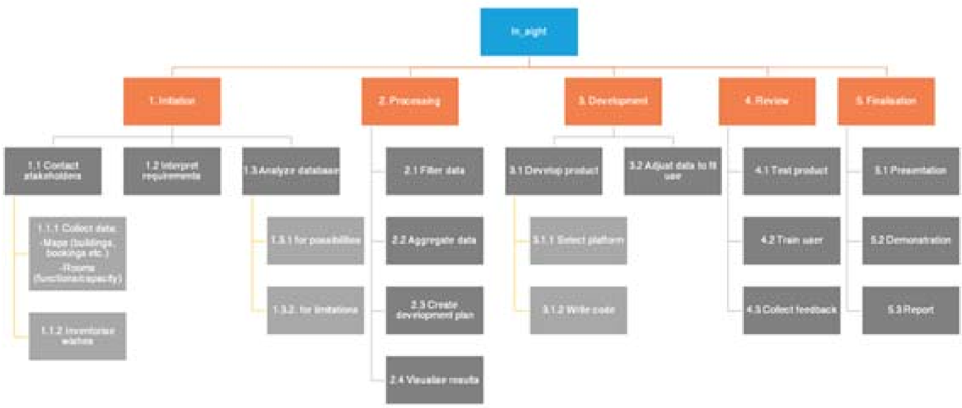


Figure VIII: Work Breakdown Structure (WBS).

I.VI. How did you deal with risks?

A risk register was created for the Mid-Term review (Table II). The register is used to identify possible risks, their causes and their possible effects on the project and project activities and to document how to handle these risks to decrease their impact. The risk responses are taken from the ATOM model, and the post response assessment is used to document whether the response decreases the probability of the risk occurring or decreases the impact of the risk if it does occur. Relevant high impact risks were communicated to both FMRE and the project mentor.

Cause	Risk event	Consequence	Project promise	Proba-bility	Impact	Description	ATOM	Proba-bility	Impact
The map of AP locations cannot be provided by the ICT specialist	**The access point (AP) locations are not available on a map**	Mandatory decrease in scope due to manual validation of AP locations	Time, Scope	HIGH	VERY HIGH	Discuss with mentor and client the issues regarding this risk towards the scope and validation of the project	Reduce	HIGH	MODE-RATE

Table V: Example of Occurred Risk.

An example of the necessity of the defined risk register is depicted in table 3. A risk anticipated was the unavailability of a map of Access Points (APs). When the risk occurred, the research plans were already defined so that the unknown of AP locations would be of minimum impact, as both the probability and the impact of the risk were anticipated as high.

I.VII. Did you change the approach during the project? Did you experiment? If not, why not? If so, how?

The initial approach was not changed during the project. As the duration of the project was relatively short, the roles within the project were fixed, so that each team member could fully understand their role. The focus on roles was also beneficial in distributing the workload and responsibilities among the team members.

I.VII.I. Team Member Evaluation

Halfway through the project a team evaluation was done in the form of a Leary Rose test. This test consisted of few series of questions about the other team members as well as the overall group functioning. The Leary Rose provided insight into the attitudes of the different team members. In total eight different attitudes could be distinguished: leading, helping, cooperative, following, withdrawn, rebellious, offensive and competitive (Figure IX).

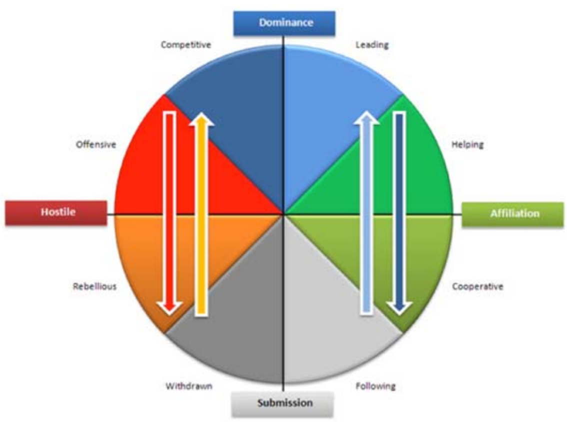


Figure IX: Overview of the different attitudes distinguished in the Leary Rose.

The outcome of the test showed that all team members act according to their defined team role, leading to a small discussion on attitude awareness. The results could also have led to reconsider the team roles. The test provides the team members with the opportunity to process feedback and improve their individual skills. An example of the Leary Rose output is shown in figure X and table VI. The main purpose of the test is to evaluate other team members anonymously, thereby being able to discuss roles and attitudes in the team on a project oriented level, rather than a personal one.

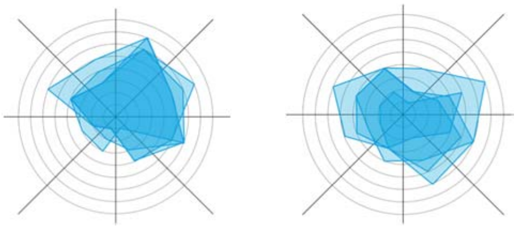


Figure X: Example of the Leary Rose outcome of team member X (left) and Y (right). As can be seen, team member X is in a leading and cooperative position, which fits the description of either a Project Manager or Sprint Master Master. Team member Y on the other hand, is not specifically captured in one area. The meaning of this outcome, rather leading and docile attitude is not desirable within a team. The identity of the team members is kept private deliberately.

Outcome Leary Rose	**Team discussion**
Team behaviour of person X fits the team role	No changes in team role
	Challenge team member to participate in another team role (initiates possible risk)
Team behaviour of person Y does not fit the team role	No changes in team role > Challenge team member to adjust behaviour in project team according to defined team role
	Change team roles > Switch roles with another team member in order to exploit individual skills better

Table VI: Overview team discussion on the basis of Leary Rose outcome.

Another test, the Team Building Monitor, was done by each team member individually. With this test, all the members had the opportunity to anonymously give scores for the team work, whereas these results might not have come forward in general meetings.

I.VII.II. Team Functioning Evaluation

In order to obtain a quick overview of the progress of the project team, all team members participated in an online 360 test. The team progress is evaluated with the teambuilding-monitor, which consists of 35 scalar questions. The test presents an overview of the team progress three main topics that capture 15 ranked aspects (Figure XI). The three main topics that are used to discuss the results are the content level, procedural level and interaction level.

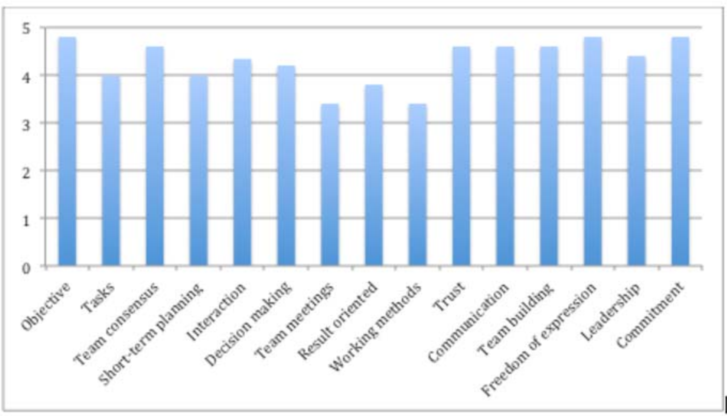


Figure XI: Overview of the results from the team building monitor.

Content level

The content level is focussed on general working methods. All team members perceive a clear objective and the task division scores also a sufficient 4 out of 5. The tasks are discussed after each meeting and noted in ASANA. This online planning tool captures all tasks and deadlines on a calendar. The tool allows the team members to tick tasks when completed and provides direct feedback and overview of each team member's progress, as well as the general process (Figure XII).

Figure XII: Progress as listed on ASANA at 14-06-2016.

The short-term planning scores sufficient, mainly due to the efficient use of tasks and deadlines in ASANA and discussions during the team meetings. The discussions in the meetings and the group dynamics (see procedural level and interaction level) grades team consensus parameter 4.5 out of 5. The project team is satisfied with the current implementation of meetings and daily updates.

Procedural level

The team meetings and working methods both score a 3.3 out of 5. This is still an adequate mark, but below the average score. In a small discussion we discovered that the meetings take too much time. To improve this process we need to discuss the results during the team meetings in a quicker way. The decision making process scored a 4.1. The discussions about decisions in the sprints work well for this group. The trust and leadership are evaluated with a 4.5 and 4.2 out of 5.

Interaction level

The interaction between the team members scores a 4.1. This is because of the 3 team meetings per week and the clear distinction of tasks. The communication scores a 4.5 out of 5. The use of the ASANA website for publishing the tasks and deadlines is a good decision. The freedom of expression scores a 4.7 out of 5. This is a really important factor. It is useful to get all differed kind of ideas from all the team members to create a (new) method. We are also happy with the commitment of 4.7. We are an enthusiastic team!

I.VIII. How did you divide the different roles in your team?

I.VIII.I. FMRE (Bart Valks)

- Commissioning party, regarded as main client for the project.
- Weekly meetings with project in_sight (email, discussion).

I.VIII.II. Mentor (Martijn Meijers)

- Commissioning party on behalf of the Master Geomatics.
- Weekly meetings with project in_sight (email, discussion).

I.VIII.III. Project Manager (Hiske Braaksma)

- Responsible to manage the overall progress of in_sight.
- Contact person for the main client (FMRE).

I.VIII.IV. KPI Specialist (Rob Braggaar)

- Responsible for systems engineering, thereby also defining the Key Performance Indicators.
- In charge of monitoring the defined indicators and alert when off-track. Mainly in contact with Project Manager.

I.VIII.V. Sprint-master (Fanny Bot)

- Responsible to manage the weekly progress of in_sight.
- The sprint-master facilitates the process and alerts the Project Manager when progress is affected. The Project Manager is also contacted when specific input from the main client (FMRE) is required.
- Contact person for the representative from Geomatics (mentor).

I.VIII.VI. Development Team (Bart Staats, Birgit Ligtvoet)

- Weekly sprints will be performed (understand, define, diverge, decide, evaluate). Therefore the Development Team has most interactions with the Scrum Master.
- Responsible for the in-depth focus of the design sprints.

I.IX. Team effort

The roles and responsibilities are clearly defined and divided between the team members. Because of the size of the group and the scale of the project, every group member will have a clear technical contribution to the project. The main goal of subdividing the tasks is to clearly state a chain of command and responsibilities for the other team members.

I.X. How can groups in the next Synthesis project benefit from your experiences? What would you have done otherwise if you could start again?

- Before the Mid-term presentations, we used some of the team performance and personality tests on 360test.nl. The Leary rose test is useful for team building and identifying whether a person has an appropriate role in the team.
- Each student should be capable of defining for himself which role would be an appropriate one. If a person is not content with their role they should speak up. It might be possible to switch team rolls around to add other activities to a team member's tasks.
- Taking on a task that might not fit a person's experience may be a good lesson for team coorperation, but in a short but influential project like this, it might be wiser to divide roles according to apparent skill sets and talents.
- Good and clear planning allows team members to be up to date on their responsibilities as well as ensuring no double work is done.
- A GANNT chart is a complicated visualisation of the planning. If all the tasks are added to the chart, the total overview is lost. If you only add general tasks the document gives not enough information. We used the work breakdown structure to define tasks in our sprints. In the sprint meetings we defined specific tasks per user which we added to ASANA. We think this is a better structure than the GANNT chart and advise the next groups to use this as well.

Appendix IV: Sprints

Paragraph 3.4.3 mentions the concept of sprints, as a framework the project planning is built upon. This appendix slightly expands on the concept of these sprints, as well as the practical outcomes. The global planning of the project can be found in table VII. The general task division can be found in the Work Breakdown Structure (WBS) in Appendix VI.

Week	1	2	3	4	5	6	7	8	9
	Phase 1	Phase 2				Phase 3			Phase 4
	Project Plan								
		Sprint 1: Campus / Faculty							
				Sprint 2: Floor					
					Report				
						Sprint 3: Floor			
							Sprint 4: Room / AP		
								Report	
									Sympo-sium

Table VII: Sprint Planning and Progress. Orange Line: Deadline. Pink Block: Delay.

I.XI. Sprint Concept

Each design sprint is divided into six stages. Each stage is defined as follows. The following paragraphs elaborate on the practical realization of each sprint, according to the described stages.

I.XI.I. Understand

What are the user needs, business needs and technology capacities?

The first part of this sprint requires advanced preparation to understand the available resources, technology capability, skills of team members and the user needs.

I.XI.II. Define

What is the key strategy and focus?

In this phase, the team starts developing their focus and strategy. This happens by defining the main focus and the desired perspective on outcomes.

I.XI.III. Diverge

How might we explore as many ideas as possible?

Often, teams choose the first ready idea to pursue. The diverge phase encourages the team to do the opposite: to first generate as many ideas as possible before they commit to the best option. In this stage, everyone is encouraged to work individually to sketch ideas.

I.XI.IV. Decide

Select the best ideas so far.

In this section it’s time for the team to review all the ideas from the diverge phase and vote for the best options. A decision table is drawn as a conclusion.

I.XI.V. Prototype

Create an artefact that allows to test the ideas with users.

Rapid prototyping allows testing out the ideas without investing too much time or resources. By doing so, it will be known which ideas fail and which have potential in an early phase.

I.XI.VI. Validate

Test the ideas with users, business stakeholders and technical experts.
The validation phase contains activities which are focussed on testing the quality of the design and the concept. There are multiple methods to test the quality of a design: by using a KPI framework, by performing manual validation and by inviting potential users to test the design.

I.XII. Sprint 1: Campus / Faculty

Goal: Evaluate the three hypotheses on SL0 (Campus) and SL1 (Faculty).

I.XII.I. Understand (27 April)

- Create understanding of subjects: Clearly define Spatial Levels (Birgit).
  - *We may be able to use movement to filter data.* ■ file: SL level description
  - *Definition of SL0-4.* ■ file: same location
  - *Define reliability issues for SL0-4.* ■ file: same location
- Create AP Overview (Fanny).
  - *Find meaning of 'Root Area' APs.* ■ named root area when attribute *maploc* is not defined. file: 1. Phase 1 - Baseline Review/SQL_Query_phase1
  - *SL0: Total number of APs.* ■ file: same location
  - *SL1: Number of APs per faculty.* ■ file: same location
- Create overview of use types for SL0-1 (Bart).
  - *General percentage of area for each use type.* Not applicable anymore, maybe phase 3
- Define user need: ■
  - What does FMRE want to see?
    - *Nominative division of exploitation for SL0-1.*
    - *Preferably campus wide.*
    - *Proven reliability percentage.*
  - What does Geomatics/mentor want to see?
    - *Rational division of exploitation for SL0-1.*
    - *Concept applied to at least part of the dataset.*
    - *Proven reliability percentage.*

I.XII.II. Define (28 April)

- Break down ideas into meaningful categories.
  - *Mind map.* ■

Figure XIII: Mindmap Sprint 1

- Create stakeholder map.
  - *Will mainly be FMRE for sprint 1.* Overview clear enough in rich picture
- Define strategies.
  - Prepare tasks for WBS 2.Processing and WBS 3.Development.
    - *Create base map (Bart).*
    - *Validate data (Fanny, Bart).*
    - *Filter data (Birgit, Bart).* Exceeds sprint duration -> GANTT
    - *Visualize data on map.* Map created, dashboard necessary for viewing: visualizing in phase 3.
    - *Create an interface (Fanny, Rob: Geoweb Assignment).* Work in progress
    - *Visualize data in the interface.* Work in progress
  - Assess Information Flows.
    - *Visit Duration (SL0): Not applicable. People travelling across campus will be counted, as they most likely have a destination on campus. Guest travellers don't influence campus occupation, but are also not logged since they don't have an Eduroam account.*
    - *Visit Duration (SL1): Logged sessions for a user within the same building are aggregated into a visit. If the visit duration is less then fifteen minutes, the user visit is classified as travel time and filtered out.*
    - *Access Point Location (SL0): The exact location of APs is not of influence on campus level.*
    - *Access Point Location (SL1): The APs are grouped by building based on AP name.*
    - *Area Covered By Access Point (SL0): Available APs are expected to cover all buildings to be assessed.*
    - *Area Covered By Access Point (SL1): Available APs per faculty are expected to cover the building completely.*
    - *Signal Strength (SL0-1): Not applicable. Derived values for these SLs are aggregated values for relatively discernible sets of APs per building.*

I.XII.III. Diverge (29 April)

- Per hypothesis, per spatial level, develop methods to determine results.
  - 10 minutes, six ideas.
  - Per table block, shortly discuss findings.
  - Decide on best ideas (1.4 Decide).

1. drawings: 2. Phase 2 - Mid Term Review/Design Sprint 1/Sprint1_diverge_exercise
2. minutes: 160502_Bart (paragraph 3)

I.XII.IV. Decide (2 May)

- Review all the ideas from the Diverge phase, vote for best options.
- Scope.
  - Decisions:

	Reliability Figures	Occupancy students vs staff	AP system
SL0 Campus	- Aggregate faculty data. - Campus wide numbers. - Need campus footprint.	~~**More information necessary.**~~	- Theoretical conclusions. - Aggregated numbers.
SL1 Faculty	- Gathering ground truth : - Counting. - Facial recognition. - Other technical methods. - Filter out traffic. - Filtering out home connections. - Filter out static devices. - Filter out outside connections, not belonging to passers. - Two buildings: - Library. - IO (new AP system).	- **More information necessary** for separation between students and staff. - Practical Information - How much space is available per person. - How many seats are available per person. - What is the desired ratio between people and space. - Technical Information: - What is the actual user / space / seat ratio.	- Signal strength map is necessary for later SL's. - Compare building with new AP system (IO) with building with old AP system.

Figure XIV: Decision Table Sprint 1

I.XII.V. Prototype (3-6 May)

- Perform.
  - Query what is needed. File: 1. Phase 1 – Base Line Review/SQL_Queries_phase1
  - Collect data. File: 15. Testing/Ground_truth_26April2016
  - Analyse data.
    - *Filtering Strategies and Queries (Bart, Birgit).* File: 9. Query – Programming data/
    - *Calculate floor area (Bart).*
  - Create overviews/graphs.
    - *Insert room information into database (Bart).* File: 10. Map Data/Data to Database
    - *Visualize AP overview (Fanny).* For later stage.
    - *Visualize filtering methods (Birgit, Bart).* Finalize sprint 3.
    - *Visualize user / space ratio (Bart).* For later stages

I.XII.VI. Validate (6 May)

- Answer hypotheses for this sprint.
  - Sub hypotheses: Yes/No.
    - *H-1: Reliability figure based on filtering quality; 80% **not** yet reached.*
    - *H-2: Occupation in user / space ratio; staff and students **cannot** be distinguished at this point.*
    - *H-3: AP system **cannot** be assessed yet.*
- What can we say about the main hypothesis?
    - *In this stage, **no** sufficient level of knowledge has been reached.*

- KPI Assessment.
  - *The data is not yet visualized.*
  - *The data is not accessible to the user.*
  - *The reliability of the data is not yet tested.*
- Cross Cutting Topics.
  - Privacy.
    - *The project deals with personal information. In regard with correct handling of this information, the rights and restrictions have been assessed and described.* File: 2. Phase 2 – Mid Term Review/Data Protection
  - Validity and Accuracy.
    - *Data can be validated by counting actual users; methods have been compared and decided on.*
    - *Accuracy can be assessed from data validation, after data has been cleaned and filtered.*
  - Representativeness.
    - *No additional conclusions can be made yet.*
  - System of APs.
    - *No additional conclusions can be made yet.*

I.XIII. Sprint 2: Floor / AP

Goal: Evaluate the three hypotheses on SL2 (Floor) and SL3 (AP).

I.XIII.I. Understand (6-9 May)

- Create understanding of subjects: What are specific implications of data on floor level?
  - *Define theoretical AP reach in 3D field.* Phase 3
  - *Test AP reach through floors (Bart, Fanny).* Case Study EWI
- Create AP Overview (Fanny). file: 2. Phase 2 - Mid Term Review/SQL_Query_Phase2
  - *SL2: Number of APs per floor. Data might need to be filtered, according to AP reach tests. Separately handle APs in staircases and atria (e.g. orange hall at architecture).*
  - *SL3: What is the benefit of creating an overview of data on this level? Some APs may already define a single room; this might be a phase 3 question.*
- Create overview of use types for SL2-3 (Fanny).
  - *General percentage of area for both students and staff, derived from database.* File: 2. Phase 2 - Mid Term Review/SQL_Query_Phase2
- Define user need (Fanny):
  - What does FMRE want to see?
    - *Publish data derived from phase 1 on dashboard.*
    - *Nominative division of exploitation for SL2.*
    - *Preferably for faculty of Industrial Design (new AP system) versus a building with the old AP system.*
    - *Proven reliability percentage.*
  - What does Geomatics/mentor want to see?
    - *Rational division of exploitation for SL2.*
    - *Concept applied to at least part of the dataset.*
    - *Proven reliability percentage.*

I.XIII.II. Define (9 May)

- Break down ideas into meaningful categories.
  - *Mind map (Birgit).*

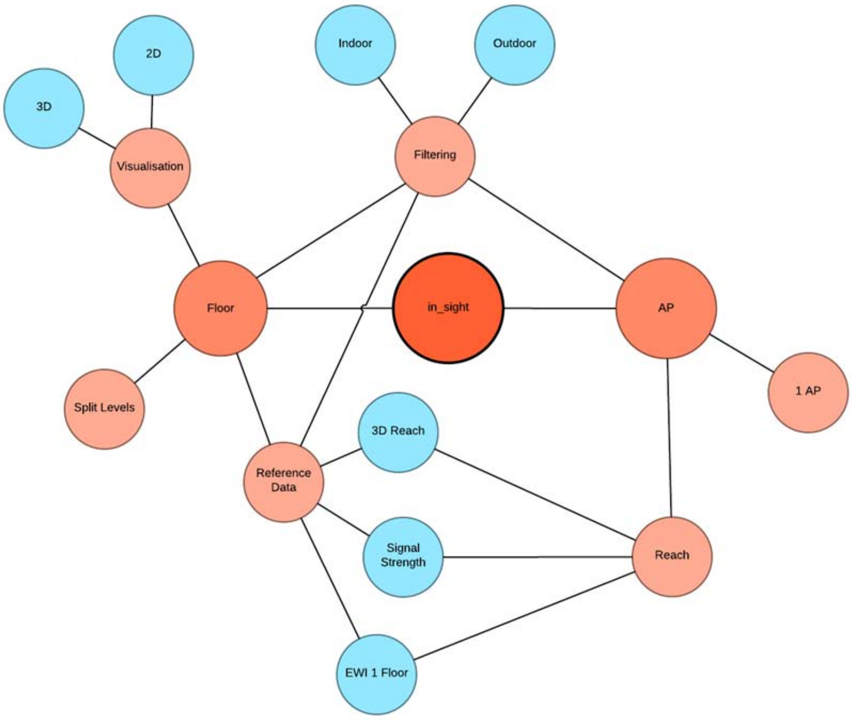


Figure XV: Mind Map Sprint 2

- Create stakeholder map.
  - *Adapt Rich Picture (Bart).*
- Define strategies.
  - Prepare tasks for WBS 2.Processing and WBS 3.Development.
    - *Filter data (Birgit, Bart).* Sprints overlay -> GANTT
    - *Create an interface (Fanny, Rob: Geoweb Assignment).*
    - *Visualize data in the interface.* Work in progress
    - *Validate data (Hiske, Bart).*
  - Assess Information Flows (Fanny, Birgit).
    - *Visit Duration (SL2): Logged sessions for a user on the same floor are aggregated into a visit. If the visit duration is less than five minutes, the user visit is classified as travel time and filtered out.*
    - *Session Duration (SL3): Every user is logged in sessions per AP. Sessions shorter than five minutes are classified as travel time and filtered out.*
    - *Access Point Location (SL2): APs are grouped per building, by floor level. Every building requires a unique handling of AP grouping per floor, as some buildings clearly log this information in apname and maploc, and others group them by several floors.*
    - *Access Point Location (SL3): As logs are registered by AP adjacency, the exact location of each AP is of maximum influence.*
    - *Area Covered By Access Point (SL2): APs might reach across floors, so signal strength may be a useful condition to filter on.*
    - *Area Covered By Access Point (SL3): A map depicting reach per AP may show whether or not the AP system covers full buildings, and therefore gives a concise representation of occupancy.*
    - *Signal Strength (SL2): Low signal strength might mean two things on this SL: the logged user is on a different floor, or the user has moved across the floor, but is still connected to a faraway AP due to roaming settings.*
    - *Signal Strength (SL3): Signal strengths indicate adjacency to AP. However, the stored signal strength values are an average value for the five minute log.*

I.XIII.III. Diverge (10 May)

- Per hypothesis, per spatial level, develop methods to determine results.
  - 10 minutes, six ideas.
  - Per table block, shortly discuss findings.
  - Decide on best ideas (2.4 Decide).

Drawings: 2. Phase 2 – Mid Term Review/Design Sprint 2/Sprint2_diverge

I.XIII.IV. Decide (10 May)

- Review all the ideas from the Diverge phase (vote for the best options. Perhaps options A, B and C can all three be tested).
- Scope.
  - Decisions:

	Reliability Figures	Occupancy students vs staff	AP system
SL2 Floor	- Collect Ground Truth: - AP Range in 3D - Count People - Two Buildings: - IO - EWI - Statistical Adaptation from Ground Truth collected at Library	- Directly from APs (EWI: 1 per floor) - Occupation in User-Space Ratio	- AP Density per M^2 - Two Buildings: - IO (new AP system) - EWI (old AP system) - Problems: - Staircases - Atria
SL3 AP	- Compare Coverage: - AP in Open Space - AP in Small Rooms - Range Map: Does not require Filters - Collect Ground Truth Derive Accuracy Rate - Statistical Analysis Improved Accuracy Rate - Use Statistical Process On Buildings to be Processed	- Classification will always have a low Accuracy - Three Cases: - Staff Use - Student Use - Mixed Use	- Visualize AP Coverage. Two cases: - Large Open Space - Small Room - Inside / Outside Ratio - Check LAN - Variable Strength Provides Classification Problems - Improvement: Max Range per AP

Figure XVI: Decision Table Sprint 2

I.XIII.V. Prototype (11-13 May)

- Perform.
  - Query what is needed ■ File: 2. Phase 2 – Mid Term Review/SQL_Queries_phase2
  - Collect data ■ *Case Studies EWI & Library*
  - Analyse data
    - *Filtering Strategies and Queries (Bart, Birgit).* File: 9. Query – Programming data/
  - Create overviews/graphs
    - *Visualize EWI data* ■ File: 15. Testing/data_ewi_en_io_12_5/map_ewi_scanners
    - *Visualize Library data* ■ File: 15. Testing/itrack validation/
    - *Create Visual Dashboard* ■ File: 16. Geoweb/user_interface_mockup

I.XIII.VI. Validate (13 May)

- Answer hypotheses for this sprint.
  - Sub hypotheses: Yes/No.
    – *H-1: Reliability figure based on filtering quality; 80%* ***not*** *yet reached.*
    – *H-2: Occupation in user / space ratio; staff and students* ***cannot*** *be distinguished at this point.*
    – *H-3: AP system provides* ***many shortcomings*** *as to assess occupancy on higher SLs.*
- What can we say about the main hypothesis?
    – *In this stage,* ***no*** *sufficient level of knowledge has been reached.*

- KPI Assessment.
    - *The data is partly visualised*
    - *The data is not yet accessible to the user.*
    - *The reliability of the data is partly tested.*
    - *Main focus on improving data processing:*
        - *Improve data filtering*
        - *Check the data acquired by third party*
    - *The platform should be adapted with user feedback to make it usable for the user*
    - *A stakeholder meeting should be planned to show preliminary results (mid-term is a suitable moment)*

- Cross Cutting Topics.
    - Privacy.
        - *The dashboard is developed on a localhost per machine, to keep the output private.*
    - Validity and Accuracy.
        - *Technical validation methods for the wifi logs are interesting to use, but difficult to reach and install, and especially expensive in time and costs.*
        - *At least for the library, the wifi logs can be validated through the itrack cameras, provided its output is reliable.*
    - Representativeness.
        - *The data is now aggregated per day, as the representativeness per hour has not been tested yet.*
        - *The data filtering processes should increase representativeness, despite the lowering of people included.*
    - System of APs.
        - *The system settings of the APs make it difficult to locate people inside the same building.*
        - *The denseness of APs makes it difficult to locate people inside the same building.*
        - *The inconsistent naming, coding and placement of APs makes it difficult to locate people inside the same building.*

I.XIV. Report

Originally, the first reporting week was planned to also focus on SL4. As one of the results of sprint 2 was that the lack of information made it difficult to define SL2, this week was redefined to include the following topics:

- Report on case studies (Bart, Hiske, Fanny).
- Report on the following subjects:
    - Scope Management (Fanny, Hiske)
    - Project Management (Hiske)
    - Risk Management (Hiske)
    - KPI Framework (Rob)
    - Filtering Methods (Birgit, Bart)
    - Dashboard (Rob, Fanny)
- Define Technical Progress

I.XV. Sprint 3: Floor / Dashboard

Goal: Evaluate the three hypotheses on SL2 (Floor) and develop the dashboard.

I.XV.I. Understand (23 May)

- Create understanding of subjects.
    - *Define theory for data separation on SL2.*
    - *Specify occupation per day.*
    - *Prepare occupation per hour.*
    - *Specify reliability calculation.*

- *Calculate exploitation: m^2 per person, percentage.*
- *Assess AP system capabilities for SL2 according to specs and measurements.*
- *Define viewpoint on students/staff separation.*

- Define global tasks.
  - *Perform case study for floor level at Architecture (Bart, Hiske).*
  - *Optional: Expand case study for floor level at EWI (Bart, Fanny).*
  - *Gather additional validation data (FMRE/other groups) (Hiske contact other groups).*
  - *Specify filtering queries for SL2 (Birgit, Bart).*
  - *Produce actual outcome (Hiske).*
  - *Publish outcome on dashboard (Geoweb: Rob, Fanny).*
  - *Visualize outcome on dashboard (Geoweb: Rob, Fanny).*
- Define specific case studies.
  - Define which building / wing etc. to process.
    - *Architecture: North West wing, all three floors plus sub floors.*
    - *EWI: 2nd floor, 6th floor.*
    - *CiTG: South partition, all six floors.*
    - *Architecture: actively measure 3D reach for single AP (Hiske).*
  - Specify (approximate) AP locations for chosen locations.
    - *Architecture: information (probably) available.*
    - *EWI: information not available.*
    - *CiTG: to be assessed.*
- Define user need:
  - What does FMRE want to see?
    - *Information on SL0-1 sufficient.*
    - *Proven reliability figures.*
    - *Define filtering in number of records / users.*
    - *Graphs: Number of users (bar) vs. exploitation percentages (line).*
  - What does Geomatics/mentor want to see?
    - *Preference for information for SL2-4.*
    - *Proven reliability figures.*

I.XV.II. Define (23 April)

- Break down ideas into meaningful categories.
  - *Mind map.*

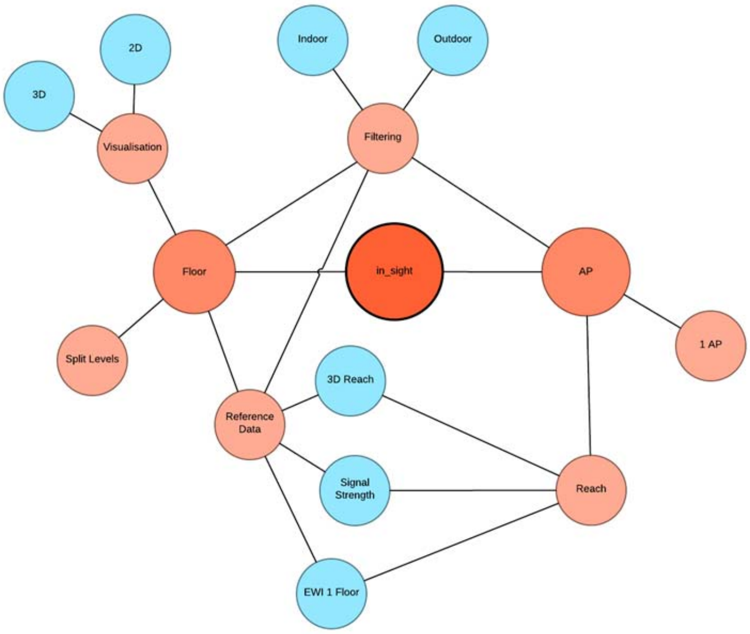


Figure XVII: Mindmap Sprint 3

- Define strategies.
  - Prepare tasks for WBS 2.Processing and WBS 3.Development.
    - *Create 3D base map.*
    - *Validate data.*
    - *Filter data.*
    - *Visualize data on map.*
    - *Visualize data in the interface.*
  - Assess Information Flows.
    - *Session Duration: Logged sessions for users on the same floor are aggregated as a visit. If the visit duration is less than ten minutes, the user visit is classified as travel time and filtered out.*
    - *AP Location: APs are grouped per building, by floor level. Every building requires a unique handling of AP grouping per floor, as some buildings clearly log this information in* apname *and* maploc.
    - *Area Covered by AP: APs might reach across floors, so signal strength may be useful to filter on.*
    - *Signal Strength: Low signal strength might mean two things: the logged user is on a different floor, or the user has moved, but is still connected to a different AP due to roaming settings. (-128??)*

I.XV.III. Diverge (25 May)

- Per hypothesis, per spatial level, develop methods to determine results.
  - 10 minutes, six ideas.
  - Per table block, shortly discuss findings.
  - Decide on best ideas (1.4 Decide).

I.XV.IV. Decide (25 May)

- Review all the ideas from the Diverge phase, vote for best options.
- Scope.
  - Decisions:

	Reliability Figures	Occupation	AP system
SL2 Floors	- Improving Factors: - Filtering - List Effect - Decreasing Factors: - Case Study effects - Effect Floors Case Study: - Three (isolated) Floors, Map Signal Strength. - EWI floor 2-6, BK OTB.	- Occupation: ppl (per floor) - Exploitation: ppl per m^2 (per floor) Task: Aggregate Excel Floorspace	- Theoretical Research on Wifi Propagation - Connection Sufficiency vs. Tracking Sufficiency Task: Check LAN! Case Study: - Single Hallway, Connect At Many Points. - EWI top floors, BK OTB

Figure XVIII: Decision Table Sprint 4

Must • SL0: o Full 3D Model (LOD1). o Interactive Graphs. • SL1: o Coloured Buildings (Exploitation). • Popup: o Building Name. o No. of Users. • Slider: o Choose Day. • Legend (Colors). • Mockups: o Colored Buildings. o Generic Footprint, Different Heights. o Colored Per Floor.	**Should** • SL1: o Bar Chart Geometry (Exploitation). o Outlines for Ideal and Footprint. • SL2: o Coloured Floors. • Popup: o Reliability Figure. • Slider: o Choose Hour.
Could • SL2: o Full Campus Application. • Slider: o Average Values for Several Days. • Export: o Export data as image. o Export data to csv. • Automatically Update Calendar Range	**Won't** • SL2: o "Scroll" Through Floors. o Heatmap (Exploitation). o OnClick: Visible Floor Separation. • SL3-4: o Publication of Data. • Slider: o Average Values for Several Hours.

Figure XIX: Table 2: MoSCoW Rules Dashboard

I.XV.V. Prototype (25-27 May)

- Perform.
  - Query what is needed.
  - Collect data.
    - *Test Case: EWI three floors, measure every few meters (Hiske, Fanny).* file: 15. Testing/ewi2/
    - *Test Case: BK three floors, measure every few meters (Hiske, Fanny).* file: 15. Testing/bk_floors
    - *Test Case: Expand EWI 1 with two subsequent floors (Hiske, Fanny).* file: 15. Testing/ewi1_io
    - *Gather additional validation data (Hiske).* file: weekend tentamen periode fellowship_legermuseum_aula. Might not be representative enough for this research.
  - Analyse data.
    - *Filtering Strategies and Queries (Bart, Birgit, Hiske).* SL1 ready, SL0 and SL2 to be finalized.
    - *Clearly define influence of filtering strategies (next sprint).*
    - *Calculate floor area (Hiske).* database
    - *Calculate occupancy as number of users (Birgit).* database

- *Calculate exploitation as floorspace per user (Birgit).* database
- *Research exploitation goals etc. in literature (Hiske)* 18 m² pp.
- *Calculate exploitation as ratio, where more floorspace per person retrieves a smaller use percentage.*
- *Specify reliability calculation (next sprint).*
- *Research theory on wifi signal propagation and AP specs (Rob).* File: 17. AP Research
- *Research theory on AP system for wifi connection vs. wifi tracking (Rob).*
- *Check LAN signal strength (Rob).* not logged

  - Create overviews/graphs.
    - *Insert room information into database (Hiske).* database
    - *Visualize test cases with measurements and AP locations (Fanny).* file: 15. Testing/Case Studies
    - *Visualize filtering methods (Birgit, Bart).* (Necessary? Next sprint)
    - *Visualize user / space ratio (Hiske).*
    - *Finalize 3D models SL0-1 on dashboard (Rob, Bart).*
    - *Research colours and floors for dashboard (Fanny, Rob, Bart).* Floors: done.
    - *Create full database – dashboard connection (Fanny, Rob, Bart).*

I.XV.VI. Validate (27 May)

- Answer hypotheses for this sprint.
  - Sub hypotheses: Yes/No.
    - *H-1: For SL2, occupation should be discernible across floors according to the* maploc *attribute in the database. However, the case studies have proven that it cannot be predicted to which AP, on which floor a user will be connected. Therefore, the occupation of floors* ***cannot*** *be determined with a sufficient reliability.*
    - *H-2: For SL2, the occupation and exploitation of floors can only be determined with very low reliability. Any further derivations cannot be accurate. The test cases are located in mixed use facilities in an attempt to gather additional information about the different use types, but since users aren't clearly distinguishable between floors, the occupation and exploitation of use types staff and students* ***cannot*** *be distinguished with any certainty.*
    - *H-3: For SL2, the AP system provides sufficient connectivity across the researched areas. However, as the system is set up to always provide coverage, it cannot be distinguished on which floor a user is. Therefore, the AP system is* ***not*** *sufficient for indicating occupation*
- What can we say about the main hypothesis?
  - *At least for SL2, the alignment of occupation and exploitation of educational facilities* ***cannot*** *be indicated with WiFi monitoring with adequate reliability.*
- KPI Assessment.
  - *The stakeholders have been contacted and their requirements have been updated.*
  - *The dashboard has seen major updates improving the user friendliness and the accessibility of the dashboard*
  - *The responsiveness of the dashboard is still a point for further improvement*
  - *The team communication scores good and the meetings are constructive*
- Cross Cutting Topics.
  - Privacy.
    - *The project deals with personal information. In regard with correct handling of this information, the rights and restrictions have been assessed and described.*
  - Validity and Accuracy.
    - *Data can be validated by counting actual users; methods have been compared and decided on.*
    - *Accuracy can be assessed from data validation, after data has been cleaned and filtered.*
  - Representativeness.
    - *On SL2, the numbers to be retrieved from the database are not sufficiently representative of real world numbers.*
  - System of APs.
    - *The system of APs does not allow for a distinction between different floors.*

I.XVI. Sprint 5: Room / Dashboard

Goal: Evaluate the three hypotheses on SL4 (Room) and develop the dashboard. As SL4 is only theoretically approachable, none of its data will be published.

I.XVI.I. Understand (30 May)

- Create understanding of subjects.
    - *Specify occupation per hour.*
    - *Specify reliability calculation.*
    - *Calculate exploitation: m^2 per person, percentage.*
    - *Assess AP system capabilities for SL2 according to specs and measurements.*
    - *Assess AP system capabilities for SL4 according to specs and measurements.*
    - *Specify why students and staff cannot be discerned.*
    - *Geoweb has to be handed in this week!!*
- Define global tasks.
    - *Gather additional validation data (FMRE/other groups) (Hiske contact other groups).*
    - *Finish filtering queries (Bart, Birgit).*
    - *Produce actual outcome (Hiske).*
    - *Publish outcome on dashboard (Geoweb: Rob, Fanny).*
    - *Visualize outcome on dashboard (Geoweb: Rob, Fanny).*
    - *Define sub questions to the main research questions, after which the final report and research structure can be modelled (All).*
- Define specific case studies.
    - Define which building / wing etc. to process.
        - *Architecture: actively measure 3D reach for single AP (Hiske, Bart).*
    - Specify (approximate) AP locations for chosen locations.
        - *Visualize test outcome, with AP locations in (heat)map.*
- Define user need:
    - What does FMRE want to see?
        - *Information on SL0-1 sufficient.*
        - *Proven reliability figures.*
        - *Define filtering in number of records / users.*
        - *Graphs: Number of users (bar) vs. exploitation percentages (line).*
    - What does Geomatics/mentor want to see?
        - *Preference for information for SL2-4.*
        - *Proven reliability figures.*
        - *Visuals: cylinders filled for exploitation for SL2.*

I.XVI.II. Define (30 May)

- Break down ideas into meaningful categories.
    - *Subquestions: will guide structure of progress and report.*
- Define strategies.
    - Prepare tasks for WBS 2.Processing and WBS 3.Development.
        - *Create 3D model SL2.*
        - *Validate data.*
        - *Filter data.*
        - *Visualize data on map.*
        - *Visualize data in the interface.*
    - Assess Information Flows.
        - *Session Duration (SL3-4): Every user is logged in sessions per AP. Sessions shorter than five minutes are classified as travel time and filtered out.*

- *Access Point Location (SL3): As logs are registered by AP adjacency, the exact location of each AP is of maximum influence.*
- *Access Point Location (SL4): As the executed case studies have proven that logs at a certain AP cannot even discern on which floor a device is located, the exact room cannot be determined with any certainty.*
- *Area Covered By Access Point (SL3): A map depicting reach per AP may show whether or not the AP system covers full buildings, and therefore gives a concise representation of occupancy.*
- *Area Covered By Access Point (SL4): As the executed case studies have proven that logs at a certain AP cannot even discern on which floor a device is located, the exact room cannot be determined with any certainty.*
- *Signal Strength (SL3): Signal strengths indicate adjacency to AP. However, the stored signal strength values are an average value for the five minute log.*
- *Signal Strength (SL4): Signal strengths indicate adjacency to AP. Though the stored signal strength values are an average value for the five minute log, they may be used to decide on a reliability figure.*

I.XVI.III. Diverge (1 June)

- Per hypothesis, per spatial level, develop methods to determine results.
  - 10 minutes, six ideas.
  - Per table block, shortly discuss findings.
  - Decide on best ideas (1.4 Decide).

I.XVI.IV. Decide (1 June)

- Review all the ideas from the Diverge phase, vote for best options.
- Scope.
  - Decisions:

	Reliability Figures	Occupation	AP system
SL4 Room	- As reliability on floor level is proven to be quite low, the same will count for room level. - No figures to be presented. - Low reliability needs to be separately proven. Case Study: - Count amount of people in several rooms around single AP to test anticipated low reliability.	- Occupation: ppl (per floor) - Exploitation: ppl per m^2 (per floor), Translated into percentage. Publish these in graphs on dashboard.	- Theoretical Research on Wifi Propagation - Connection Sufficiency vs. Tracking Sufficiency Case Study: - Single AP, map signal strength from different rooms, across different floors.

Figure XX: Decision Table Sprint 5

Must	**Should**
• SL0: o Full 3D Model (LOD1). o Interactive Graphs. • SL1: o Coloured Buildings (Exploitation). • Popup: o Building Name. o No. of Users. • Slider: o Choose Day. • Legend (Colors). • Mockups: o Colored Buildings. o Generic Footprint, Different Heights. o Coloured Per Floor.	• SL1: o Bar Chart Geometry (Exploitation). o Outlines for Ideal and Footprint. • SL2: o Coloured Floors. • Popup: o Reliability Figure. • Slider: o Choose Hour.

Could	Won't
• SL2: o Full Campus Application. • Slider: o Average Values for Several Days. • Export: o Export data as image. o Export data to csv. • Automatically Update Calendar Range	• SL2: o "Scroll" Through Floors. o Heat map (Exploitation). o OnClick: Visible Floor Separation. • SL3-4: o Publication of Data. • Slider: o Average Values for Several Hours.

Figure XXI: Table 4: Progress on MoSCoW Rules Dashboard

I.XVI.V. Prototype (25-27 May)

- Perform.
  - Query what is needed.
    - *Specify Occupation per hour (Birgit).*
    - *Calculate exploitation:* m^2 *per person, percentage (Hiske).*
    - *Specify why students and staff cannot be distinguished.*
  - Collect data.
    - *Test Case: BK single AP, measure rssi (Hiske, Bart, Fanny).* file: 15. Testing/bk ap
    - *Test Case: Count no of people in room (Fanny (?)).* File: 15. Testing/GroundTruth.xls
    - *Gather additional validation data (Hiske).* file: 15. Testing/////////
  - Analyse data.
    - *Filtering Strategies and Queries (Bart, Birgit, Hiske).* SL1 ready, SL0 and SL2 to be finalized.
    - *Clearly define influence of filtering strategies (Birgit).*
    - *Calculate exploitation as ratio, where more floorspace per person retrieves a smaller use percentage.*
    - *Specify reliability calculation.* May not be expressable in numbers.
    - *Research theory on wifi signal propagation and AP specs (Rob).* File: 17. AP Research
    - *Research theory on AP system for wifi connection vs. wifi tracking (Rob).* File: 17. AP Research.
  - Create overviews/graphs.
    - *Visualize filtering methods (Birgit, Bart).*
    - *Visualize user / space ratio (Hiske).*
    - *Create full database – dashboard connection (Fanny, Rob, Bart).*
    - *Publish data on dashboard for SL0-2 (Fanny, Rob, Bart).*

I.XVI.VI. Validate (27 May)

- Answer hypotheses for this sprint.
  - Sub hypotheses: Yes/No.
    - *H-1: For SL4, the distinction between different rooms is generally not made in the database, except for some larger lecture halls. The performed test cases indicate that this distinction cannot be made based on* rssi. *Therefore, the occupation of rooms* ***cannot*** *be determined with 80% or more reliability.*
    - *H-2: For SL4, the occupation and exploitation of floors can only be determined with very low reliability. Any further derivations cannot be accurate. The test cases for higher levels are located in mixed use facilities in an attempt to gather additional information about the different use types, but since users aren't clearly distinguishable between floors, the occupation and exploitation of use types staff and students* ***cannot*** *be distinguished with any certainty.*
    - *H-3: For SL4, the AP system provides sufficient connectivity across the researched areas. However, as the system is set up to always provide coverage, it cannot be distinguished on which floor a user is. Therefore, the AP system is* ***not*** *sufficient for indicating occupation*
- What can we say about the main hypothesis?
  - *At least for 4 the alignment of occupation and exploitation of educational facilities* ***cannot*** *be indicated with WiFi monitoring with adequate reliability.*

- KPI Assessment.
    - *The responsiveness of the dashboard is improved and is now (near)instant*
    - *Feedback from stakeholders has been processed, which led to the addition of some extra input functions*
    - *Documentation has been updated and extended so that the user is able to run the application on his own computer*
    - *Data is further refined for use in the dashboard.*
    - *Visualisation is improved for SL1 and SL2*
- Cross Cutting Topics.
    - Privacy.
        - *The project deals with personal information. In regard with correct handling of this information, the rights and restrictions have been assessed and described.*
    - Validity and Accuracy.
        - *Data can be validated by counting actual users; methods have been compared and decided on.*
        - *Accuracy can be assessed from data validation, after data has been cleaned and filtered.*
    - Representativeness.
        - *On SL2, the numbers to be retrieved from the database are not sufficiently representative of real world numbers.*
    - System of APs.
        - *The system of APs does not allow for a distinction between different floors.*

I.XVII. Report

During the second report week, the dashboard is finished and the final technical review is prepared. All tasks from former sprints that have suffered delay are finished, so that the project can be finalized.

Must
- SL0:
    - Full 3D Model (LOD1).
    - Interactive Graphs.
- SL1:
    - Coloured Buildings (Exploitation).
- Popup:
    - Building Name.
    - No. of Users.
- Slider:
    - Choose Day.
- Legend (Colours).
- Mock-ups:
    - Coloured Buildings.
    - Generic Footprint, Different Heights.
    - Coloured Per Floor.

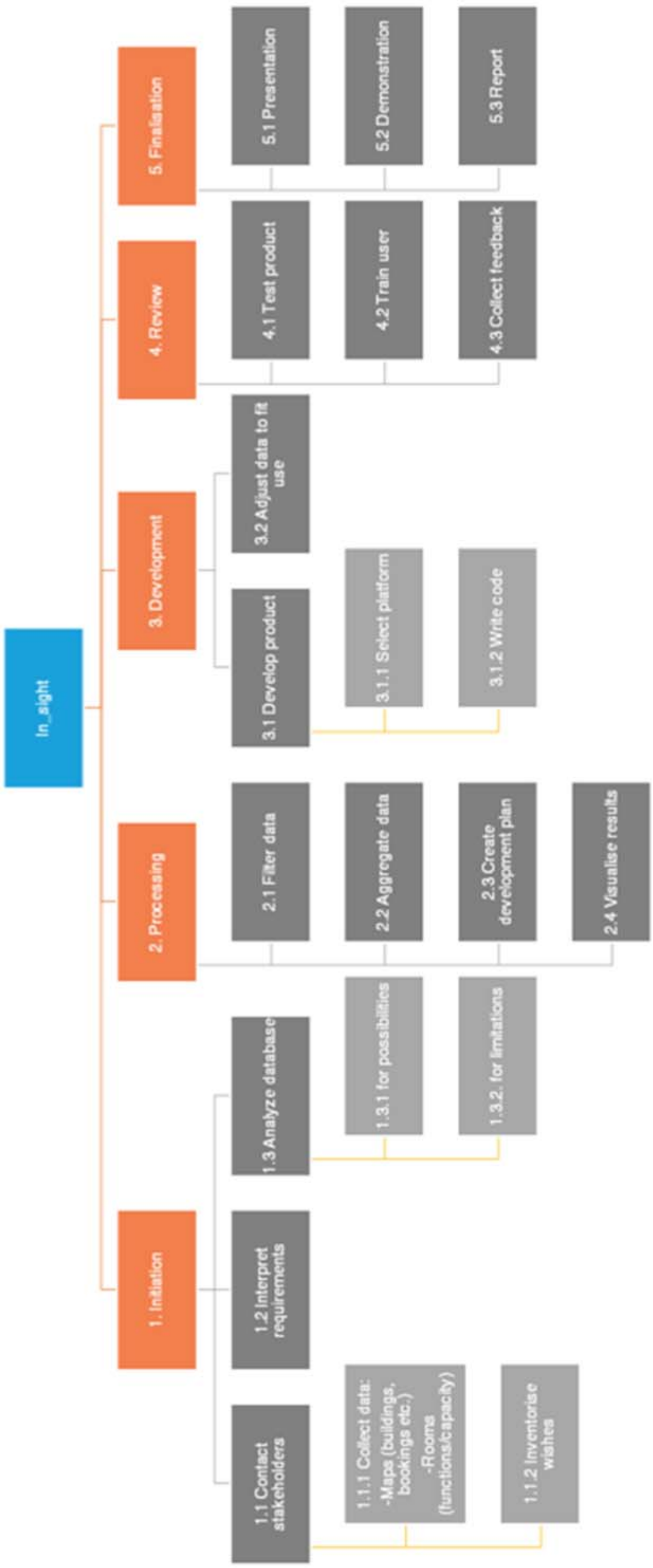
Should
- SL1:
    - Bar Chart Geometry (Exploitation).
    - Outlines for Ideal and Footprint.
- SL2:
    - Coloured Floors.
- Popup:
    - Reliability Figure.
- Slider:
    - Choose Hour.

Could
- SL2:
    - Full Campus Application.
- Slider:
    - Average Values for Several Days.
- Export:
    - Export data as image.
    - Export data to csv.
- Automatically Update Calendar Range

Won't
- SL2:
    - "Scroll" Through Floors.
    - Heat map (Exploitation).
    - On Click: Visible Floor Separation.
- SL3-4:
    - Publication of Data.
- Slider:
    - Average Values for Several Hours.

Figure XXII: Progress on MoSCoW Rules Dashboard

Appendix V: Work Breakdown Structure (WBS)



Appendix VI: Filtering

I.XVIII. Query 1: Identifying Static Device Users

```
create materialized view g01_static_devices as
select username, round(sum(macs_each_day)/20) as avg_macs, max(macs_each_day) as max_macs
from(
        select username, count(distinct mac) as macs_each_day, asstime::date as datum
        from wifilog
        where asstime between current_date -30 and current_date
        group by username, asstime::date
        ) as t1
group by username
having round(sum(macs_each_day)/20) >3 and  max(macs_each_day) >4;
```

I.XIX. Query 2-4: Identifying Users Living on Campus

I.XIX.I. Query 2

```
create materialized view g01_user_freq_night as
select username
from(   select username, count(distinct asstime::date) as user_freq
                from(   select username,apname, asstime,sesdur
                                from wifilog
                                where asstime between current_date-30 and current_date
                                and sesdur > interval '10 minutes'
                        )       as t1
                where   asstime::time between time '23:00' and '23:59'
                        or
                        asstime::time between time '00:00' and '06:00'
                        or
                        (asstime+sesdur)::time between time '23:00' and '23:59'
                        or
                        (asstime+sesdur)::time between time '00:00' and '06:00'
                group by username
                ) as t2
where user_freq >= 10
group by username;
```

I.XIX.II. Query 3

```
create materialized view g01_user_fav_ap as
select  n1.username, apname
from(   select username, apname, sum(sesdur)as total_con
        from wifilog
        where
                asstime::time between time '23:00' and '23:59'
                or
                asstime::time between time '00:00' and '06:00'
                or
                (asstime+sesdur)::time between time '23:00' and '23:59'
                or
                (asstime+sesdur)::time between time '00:00' and '06:00'
        group by username, apname
        ) as n1

        join(   select username, max(sum) as max_con
                from(   select username, apname, sum(sesdur)
                        from wifilog
                        where
                                asstime::time between time '23:00' and '23:59'
                                or
                                asstime::time between time '00:00' and '06:00'
```

```
                    or
                    (asstime+sesdur)::time between time '23:00' and '23:59'
                    or
                    (asstime+sesdur)::time between time '00:00' and '06:00'
                group by username, apname
                ) as n1
            group by username
            ) as n2
        on n1.username = n2.username
where   total_con = max_con;
```

I.XIX.III. Query 4

```
create materialized view g01_users_living_oncamp as
        select t1.username, t1.apname
        from(
            select g01_user_fav_ap.username,
            g01_user_fav_ap.apname
            from g01_user_fav_ap
            join g01_user_freq_night ON g01_user_fav_ap.username =
            g01_user_freq_night.username
            ) as t1
        where t1.username not in(
            select username
            from g01_static_devices);
```

I.XX. Query 5: Visit Creation and Filtering of Users Living on Campus for SL1

```
create materialized view g01_sl1_vnol as
select username,building_no,sum(true_sesdur) as fac_visdur, min(asstime) as start_t
from(
        select  username, building_no, true_sesdur, asstime, sum(changes::int) over (order by
        username,asstime)
        from(
            select   username,
            building_no,
            asstime,
            case     when lead(username) over w1=username and asstime+ sesdur >
                     lead(asstime)over w1
                     then lead(asstime) over w1 -asstime
                     when lead(asstime) over w1 = null
                     then sesdur
                     else sesdur
            end as true_sesdur,
            lag(building_no) over (order by username,asstime) is distinct from building_no
            as changes
            from     (
                     select username,
                     case     when char_length(split_part(apname, '-',2)) = 0
                              then 999
                              when split_part(apname, '-',2) = '134'
                              then 34
                                       when split_part(apname, '-',2) = '132'
                              then 32
                              else split_part(apname, '-',2)::int

                     end as building_no,
                     asstime,
                     sesdur
                     from wifilog
                     where    not exists(
                                       select 1
                                       from g01_processed
                                       where wifilog.asstime::date = g01_processed.datum)
```

```
				and not exists(
						select 1
						from g01_users_living_oncamp
						where wifilog.username =
						g01_users_living_oncamp.username
						and wifilog.apname = g01_users_living_oncamp.apname
						)
				order by username, asstime
			) as t1
			window w1 as (partition by username order by username, asstime,building_no)
			) as t2
		)as t3
group by username,sum, building_no, asstime::date;

create index vnol_username
on g01_sl1_vnol
using btree(username);

create index vnol_date
on g01_sl1_vnol
using btree(start_t);

create index vnol_visdur
on g01_sl1_vnol
using btree(fac_visdur);

create index vnol_building
on g01_sl1_vnol
using btree(building_no);
```

I.XXI. Query 6: Calculating Occupation and Exploitation for SL1

```
create materialized view g01_sl1_moeh as
select building_no as buildingid, datum, l_hour,
	case when visitors is null
		then 0
		else round(visitors)
	end as occupation,
	case when visitors != 0
		then round(faculty_nuttigo/visitors)
		else faculty_nuttigo
	end as exploitation
from(
	select building_no, datum, l_hour, extract(epoch from sum(case_1))/ extract(epoch from interval
	'1 hour') as visitors
	from(	select l_hour::time
		from generate_series(	'2016-01-01'::timestamp,
					'2016-01-01 23:00:00'::timestamp,
					'1 hour') as l_hour
		)as t2,
	lateral(	select	building_no,
			start_t::date as datum,
			case	when start_t::time between l_hour and l_hour + interval '3599 seconds'
				and (start_t::time + fac_visdur) > (l_hour + interval '3599 seconds')
				then l_hour + interval '3599 seconds' - start_t::time
				when start_t::time < l_hour and (start_t::time+fac_visdur) > l_hour +
				interval '3599 seconds'
				then interval '3599 seconds'
				when start_t::time < l_hour and (start_t::time + fac_visdur) between
				l_hour and l_hour + interval '3599 seconds'
				then (start_t::time + fac_visdur)- l_hour
				when start_t::time between l_hour and l_hour + interval '3599 seconds'
				and (start_t::time +fac_visdur) < l_hour + interval '3599 seconds'
				then  fac_visdur
				else null
			end as case_1
```

```
                        from(   select g01_sl1_vnol.username, building_no, fac_visdur, start_t
                                from g01_sl1_vnol
                                left join g01_static_devices
                                on g01_sl1_vnol.username = g01_static_devices.username
                                where g01_static_devices.username is null and fac_visdur > '15
                                minutes'
                                )as t1
                        ) as c1
        group by building_no, datum, l_hour
        )as t3
inner join g01_faculty_nuttigo on t3.building_no = g01_faculty_nuttigo.buildingid;
```

I.XXII. Query 7: Calculating occupation and exploitation for SL0

```
create materialized view g01_sl1_moeh as
select building_no as buildingid, datum, l_hour,
        case when visitors is null
                then 0
                else round(visitors)
        end as occupation,
        case when visitors != 0
                then round(faculty_nuttigo/visitors)
                else faculty_nuttigo
        end as exploitation
from(
        select building_no, datum, l_hour, extract(epoch from sum(case_1))/ extract(epoch from interval
        '1 hour') as visitors
        from(   select l_hour::time
                from generate_series(   '2016-01-01'::timestamp,
                                        '2016-01-01 23:00:00'::timestamp,
                                        '1 hour') as l_hour
                )as t2,
        lateral(        select  building_no,
                        start_t::date as datum,
                        case    when start_t::time between l_hour and l_hour + interval '3599 seconds'
                                and (start_t::time + fac_visdur) > (l_hour + interval '3599 seconds')
                                then l_hour + interval '3599 seconds' - start_t::time
                                when start_t::time < l_hour and (start_t::time+fac_visdur) > l_hour +
                                interval '3599 seconds'
                                then interval '3599 seconds'
                                when start_t::time < l_hour and (start_t::time + fac_visdur) between
                                l_hour and l_hour + interval '3599 seconds'
                                then (start_t::time + fac_visdur)- l_hour
                                when start_t::time between l_hour and l_hour + interval '3599 seconds'
                                and (start_t::time +fac_visdur) < l_hour + interval '3599 seconds'
                                then  fac_visdur
                                else null
                        end as case_1
                        from(   select g01_sl1_vnol.username, building_no, fac_visdur, start_t
                                from g01_sl1_vnol
                                left join g01_static_devices
                                on g01_sl1_vnol.username = g01_static_devices.username
                                where g01_static_devices.username is null and fac_visdur > '15
                                minutes'
                                )as t1
                        ) as c1
        group by building_no, datum, l_hour
        )as t3
inner join g01_faculty_nuttigo on t3.building_no = g01_faculty_nuttigo.buildingid;
```

I.XXIII. Query 8: Visit Creation and Filtering of Users Living on Campus for SL2

```
create materialized view g01_sl2_vnol as
 select t3.username,
```

```
 t3.building_no,
 t3.floor,
 sum(t3.true_sesdur) as floor_visdur,
 min(t3.asstime) as start_t
from ( select t2.username,
        t2.building_no,
        t2.floor,
        t2.true_sesdur,
        t2.asstime,
        sum(t2.changes::integer) over (order by t2.username, t2.asstime) as sum
       from ( select t1.username,
               t1.building_no,
               t1.floor,
               t1.asstime,
                  case      when lead(t1.username) over w1 = t1.username and (t1.asstime + t1.sesdur) >
                            lead(t1.asstime) over w1
                            then lead(t1.asstime) over w1 - t1.asstime
                    when lead(t1.asstime) over w1 = null::timestamp without time zone
                            then t1.sesdur
                    else t1.sesdur
                    end as true_sesdur,
                    lag(t1.floor) over (order by t1.username, t1.asstime) is distinct from t1.floor as changes
                    from (   select wifilog.username,
                           case      when char_length(split_part(wifilog.apname, '-'::text, 2)) = 0
                                     then 999
                                     when split_part(apname, '-',2) = '134'
                                     then 34
                                     when split_part(apname, '-',2) = '132'
                                     then 32
                                     else split_part(wifilog.apname, '-'::text, 2)::integer
                              end as building_no,
                           case      when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%begane%'::text
                                     and lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                                     then 'begane grond'::text
                                     when (lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%1%diep%'::text
                                     or lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%!%'::text) and
                                     lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                                     then '1e verdieping'::text
                                     when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%2%diep%'::text
                                     and lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                                     then '2e verdieping'::text
                                     when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%3%diep%'::text
                                     and lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                                     then '3e verdieping'::text
                                     when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%4%diep%'::text
                                     and lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                                     then '4e verdieping'::text
                                     when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%5%diep%'::text
                                     and lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                                     then '5e verdieping'::text
                                     else "substring"(lower(split_part(wifilog.maploc, '>'::text, 3)), 2)
                              end as floor,
                    wifilog.asstime,
                    wifilog.sesdur
                    from wifilog
                    where
                                     not exists( select 1
                                                 from g01_processed
                                                 where wifilog.asstime::date = g01_processed.datum)
                                     and
                                     not (exists ( select 1
                                                  from g01_users_living_oncamp
                                                  where wifilog.username =
                                                  g01_users_living_oncamp.username and wifilog.apname
                                                  = g01_users_living_oncamp.apname))
                    order by wifilog.username, wifilog.asstime
                            ) t1
```

```
                    window w1 as (partition by t1.username order by t1.username, t1.asstime, t1.building_no)
                    ) t2
) t3
group by t3.username, t3.sum, t3.building_no, t3.floor, t3.asstime::date;

create index vnol2_username
on g01_sl2_vnol
using btree(username);

create index vnol2_date
on g01_sl2_vnol
using btree(start_t);

create index vnol2_visdur
on g01_sl2_vnol
using btree(floor_visdur);

create index vnol2_building
on g01_sl2_vnol
using btree(building_no);
```

I.XXIV. Query 9: Calculating occupation and exploitation for SL2

```
create materialized view g01_sl2_vnol as
 select t3.username,
    t3.building_no,
    t3.floor,
    sum(t3.true_sesdur) as floor_visdur,
    min(t3.asstime) as start_t
   from ( select t2.username,
            t2.building_no,
            t2.floor,
            t2.true_sesdur,
            t2.asstime,
            sum(t2.changes::integer) over (order by t2.username, t2.asstime) as sum
           from ( select t1.username,
                    t1.building_no,
                    t1.floor,
                    t1.asstime,
                    case        when lead(t1.username) over w1 = t1.username and (t1.asstime + t1.sesdur) >
                                lead(t1.asstime) over w1
                                then lead(t1.asstime) over w1 - t1.asstime
                                when lead(t1.asstime) over w1 = null::timestamp without time zone
                                then t1.sesdur
                        else t1.sesdur
                    end as true_sesdur,
                    lag(t1.floor) over (order by t1.username, t1.asstime) is distinct from t1.floor as changes
                   from ( select wifilog.username,
                            case    when char_length(split_part(wifilog.apname, '-'::text, 2)) = 0
                                    then 999
                                    when split_part(apname, '-',2) = '134'
                                    then 34
                                    when split_part(apname, '-',2) = '132'
                                    then 32
                                    else split_part(wifilog.apname, '-'::text, 2)::integer
                              end as building_no,
                             case when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%begane%'::text and
                                    lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                                    then 'begane grond'::text
                                    when (lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%1%diep%'::text or
                                    lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%!%'::text) and
                                    lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                                    then '1e verdieping'::text
                                    when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%2%diep%'::text and
                                    lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                                    then '2e verdieping'::text
```

```
                    when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%3%diep%'::text and
                    lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                    '3e verdieping'::text
                    when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%4%diep%'::text and
                    lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                    then '4e verdieping'::text
                    when lower(split_part(wifilog.maploc, '>'::text, 3)) ~~ '%5%diep%'::text and
                    lower(split_part(wifilog.maploc, '>'::text, 3)) not like '%-%-%'::text
                    then '5e verdieping'::text
                    else "substring"(lower(split_part(wifilog.maploc, '>'::text, 3)), 2)
                end as floor,
                wifilog.asstime,
                wifilog.sesdur
                from wifilog
                where not exists(       select 1
                                        from g01_processed
                                        where wifilog.asstime::date = g01_processed.datum)
                  and  not (exists (    select 1
                                        from g01_users_living_oncamp
                                        where wifilog.username =
                                        g01_users_living_oncamp.username and wifilog.apname =
                                        g01_users_living_oncamp.apname))
                order by wifilog.username, wifilog.asstime
                  ) t1
        window w1 as (partition by t1.username order by t1.username, t1.asstime, t1.building_no)
        ) t2
) t3
group by t3.username, t3.sum, t3.building_no, t3.floor, t3.asstime::date;

create index vnol2_username
on g01_sl2_vnol
using btree(username);

create index vnol2_date
on g01_sl2_vnol
using btree(start_t);

create index vnol2_visdur
on g01_sl2_vnol
using btree(floor_visdur);

create index vnol2_building
on g01_sl2_vnol
using btree(building_no);
```

Appendix VII: Key Performance Indicator (KPI) Framework

From the activities and product elements the most critical tasks are chosen, the indicators are listed in the tables below. These tasks will be monitored according to the KPIs listed in Tables VIII and IX. After each sprint, both indicator types are assessed and scored on a range from one to five. The argumentation for scoring and elaboration of each indicator can be found after the overview tables. The output is visualized in radar charts, as done in later paragraphs.

Indicator	Description	Method / tools
Reference data	Quality of the chosen location for the gathering of reference data.	Field testing
Data acquirement	Quality of the measurements acquired from the access points.	Desk research
Data processing	Quality of the processing applied to the data (i.e. filters, aggregation etc.)	Comparing
Product Development	Process of product development.	Discussion
Project Planning	Planning of the project; tasks versus deadlines.	GANTT
Team Communication	The communication between the team members.	Team meeting / 360 test
Client Communication	How is the communication with the stakeholders?	Check

Table VIII: Activity indicators

Indicator	Description	Method
Stakeholder requirements	Extent to which the stakeholder requirements are fulfilled.	MoSCoW rules
Reference data	Extent to how closely the measurements resemble real world numbers.	Field testing
Accessibility	The accessibility of the platform to the user.	User testing
Usability	Extent to which the platform supports the user in making decisions.	Interviews/interactive testing /survey
Documentation	Documentation of the process.	Reports
Response time	The response time of the platform when a command is issued.	Check
Availability	The availability of the online platform or source code to the user in its most recent form.	Testing
Findability	The findability of the online platform through general search engines in case of publishing the final product.	Testing

Table IX: 5.3.2. Product indicators

I.XXV. Activity Indicators

I.XXV.I. Reference Data

To validate the results it is necessary to collect reference data. The score for the reference data indicator indicates how good the chosen sampling locations are chosen, i.e. represent the real world values. A score of 1 indicates that the location is chosen randomly. A score of 5 indicates that the reference data is collected at both a bottleneck, single entrance spot (3) as well as a location that provides other methods of validating as well, e.g. camera visitor counting.

I.XXV.II. Data Acquirement

The values that are obtained during the project are coming directly from wifi access points. This indicator is about the accuracy of the logging of the access points. A score of 1 indicates that no quality information about the system is known or that the system is offline. A score of 3 indicates that everyone is able to connect. A score of 5 also means that the location of the access points is known. The location of the access can be used to determine the approximate location of the person connected to it (+- 20 meters).

I.XXV.III. Data Processing

The raw data that is collected during the project needs to be processed. One person can carry multiple devices that are all connected through a single Eduroam user account. There are also static devices that are not counted towards the occupancy numbers. Lastly, travelling people have to be filtered out as they are not present in the room, but are just connecting while they are passing a nearby access point. 1 Indicates that the raw data is used. A score of 5 indicates that both the static devices are filtered (3) as well as travelling people (4) and people sitting just outside the building (5).

I.XXV.IV. Product Development

The stakeholders in this project are interested in the end product besides the research done. The development of the products is monitored with a selection of tools. This indicator measures how well the organization within the team aligns with the envisioned product or goals. For this indicator we used discussions within the team meetings to gauge the process. These discussions are documented in the project minutes. A score of 1 means that the documentation is lacking. A score of 5 means that all tasks are clearly allocated to different team members (3) and documented.

I.XXV.V. Project Planning

Besides the process organization we also monitor the planning. To stay on track with the project we use different planning tools: minute tasks, a GANTT chart and an online planning tool (Asana)(Figure 1). The tasks of the GANTT chart are duplicated to the online planning tool where we can easily monitor the project status. 1 Means the project is not on track or that no information on the status of the project is available. 3 Indicates that the project is on track according to the progress tracker, but that not all tasks are entered. A score of 5 means the project is on track and that all tasks are entered in the planning tools.

Figure XXIII: Progress monitor

I.XXV.VI. Team Communication

Teamwork is very important during the synthesis project. The project manager plays an important role in dividing the workload among the team members and the communication between team members. For this indicator we use online personality tests. Before taking the test we sketch expected personality profiles of each team member. Afterwards we check if everyone in the correct position according to the outcome profile. A score of 1 indicates that all of the team members show large discrepancies between the expected profile and the actual profile. 5 Means that the discrepancies are very small or none.

I.XXV.VII. Client Communication

During the project the communication with the different stakeholders is important so that the requirements can be inventoried at the start of the project or updated during the project. A score of 1 is given if there is online one meeting between the project group and the stakeholders. A score of 3 is given when more than one stakeholder meeting is done. A score of 5 means that a meeting is arranged every sprint and that intermediary results are discussed.

I.XXVI. Product Indicators

I.XXVI.I. Stakeholder Requirements

This indicator is related to the client communication, but this time focussed on the developed products. We measure the requirements with MoSCoW rules. All the Must requirements should be done (3). If the Should requirements are fulfilled at the stage of the sprint a score of 4 is given. A score of 5 is given when the could requirements are also implemented.

I.XXVI.II. Reference Data

This time the reference data indicator is focussed towards the collected data (the product) instead of the activity of collecting the data. We can compare the numbers we have collected with numbers acquired from the AP measurements and other visitor counting techniques such as camera tracking. If we cannot say anything useful about the quality of our measurements with the use of reference data a score of 1 is given. A score of 3 is given when the reference data is consistent for multiple moments. A score of 5 indicates that the reference data is consistent and that captured for multiple locations.

I.XXVI.III. Accessibility

For the online platform to be useful to the end user, FMRE, it is important that he can access the platform from his own computer. 1 Means this is not possible. A score of 3 is given when the user can use the software on his own, but that he has to install software in order to use it. A score of 5 means that the user can access the program without installing any software, from any computer.

I.XXVI.IV. Usability

The usability is the impact of the product, the degree of which the product can aid the user to take decisions. This indicator is measured with user surveys and during interactive testing. A score from 1-5 can be given by the user.

I.XXVI.V. Documentation

Documentation can be used by the user to find out about the workings of the product. It can also provide clues to how the research behind the product is conducted and provide details on the technical quality.

1 If the documentation is not available to the user. 5 Means the documentation is according to the project guide (3) and available for all stakeholders (5).

I.XXVI.VI. Response time

When using the online platform (the end product) the response time is an important measure in the usefulness of the product. 1 Means the queries behind the visualisation take longer than 5 minutes. 5 Indicates a near real-time visualisation or a very low response time (seconds).

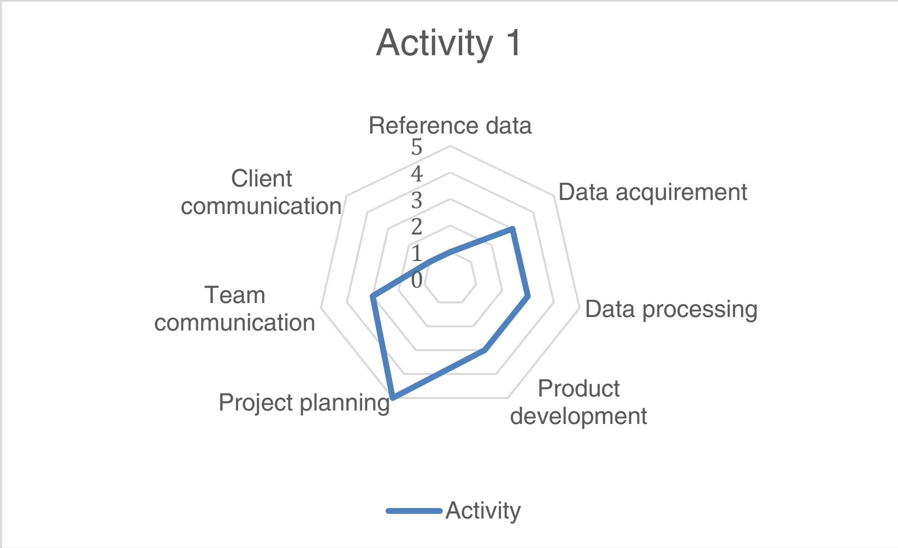
I.XXVI.VII. Availability

Availability indicates if the platform is actually available for the user. 1 If the platform is not available to the user and 5 if the platform is available.

I.XXVI.VIII. Findability

To be able to get the information a user wants, he has to be able to find the information. In case the online platform gets published on the worldwide web, this indicator measures how well the platform can be found using general search terms through an un-profiled search engine (IxQuick).

I.XXVII. Results



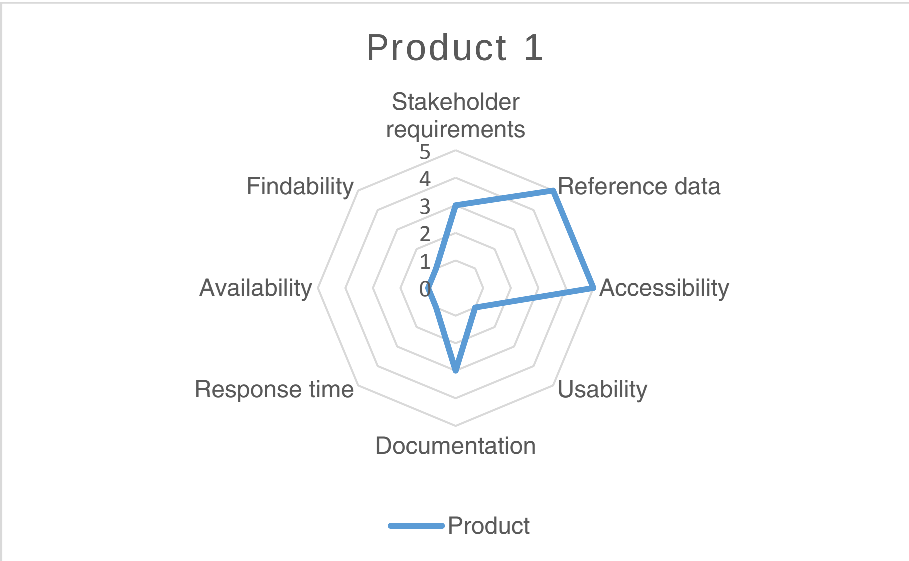
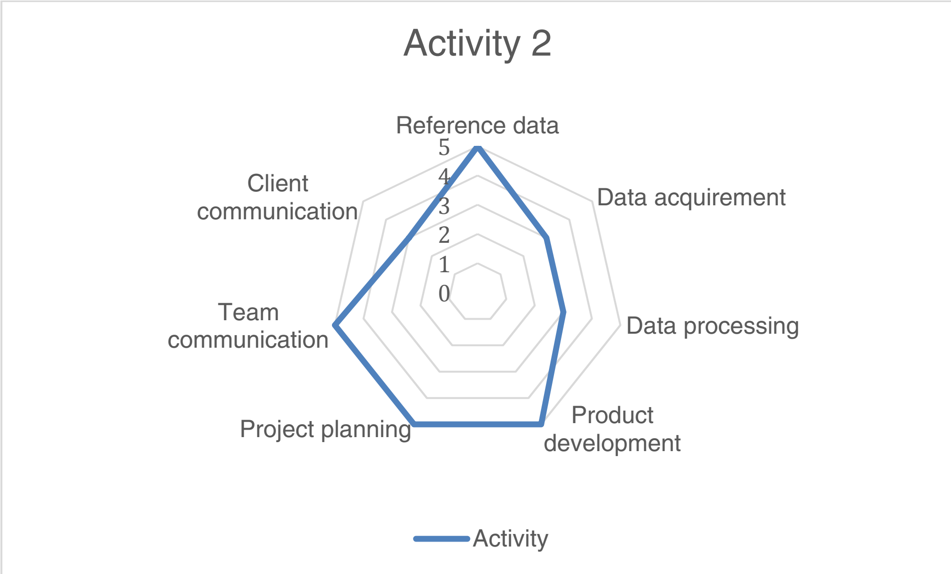


Figure XXIV: Scores sprint 1



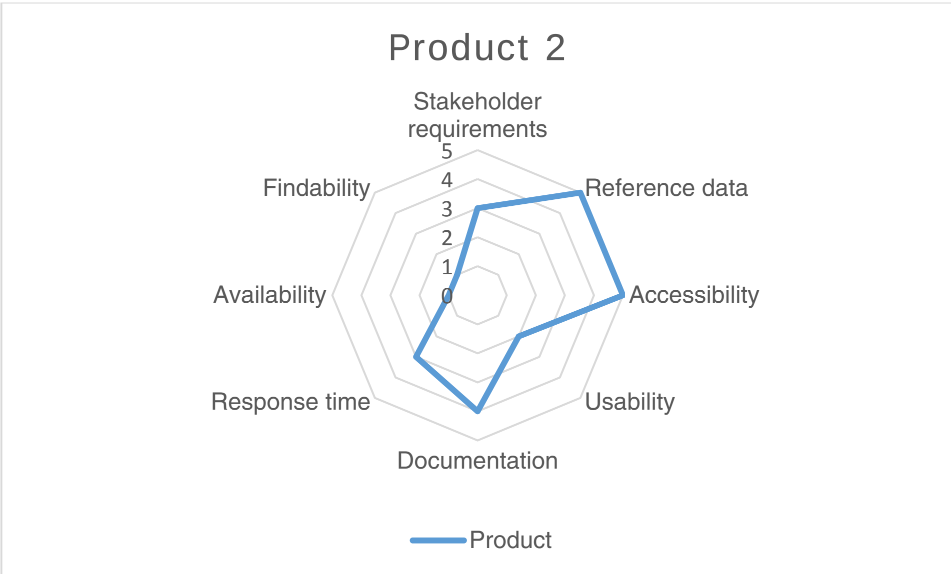
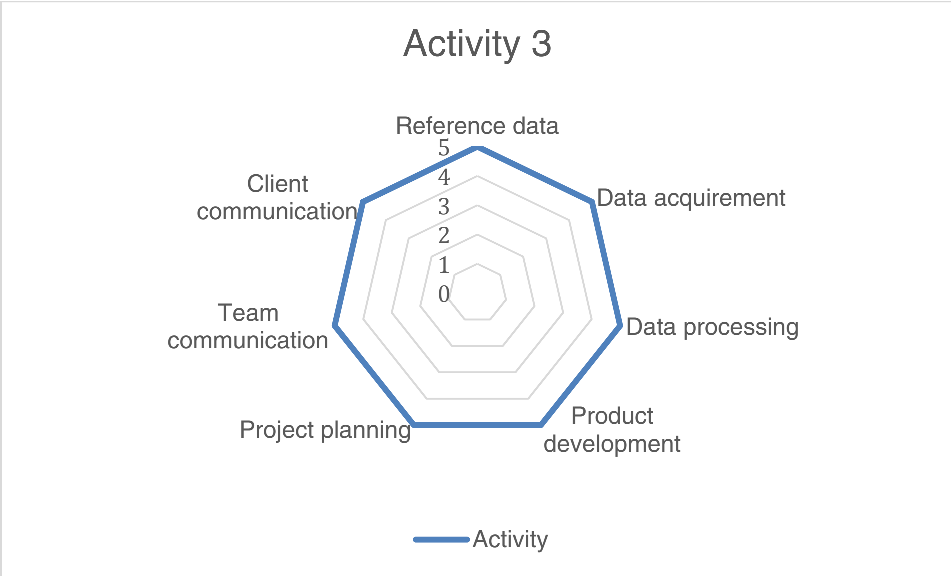


Figure XXVI: Scores sprint 3



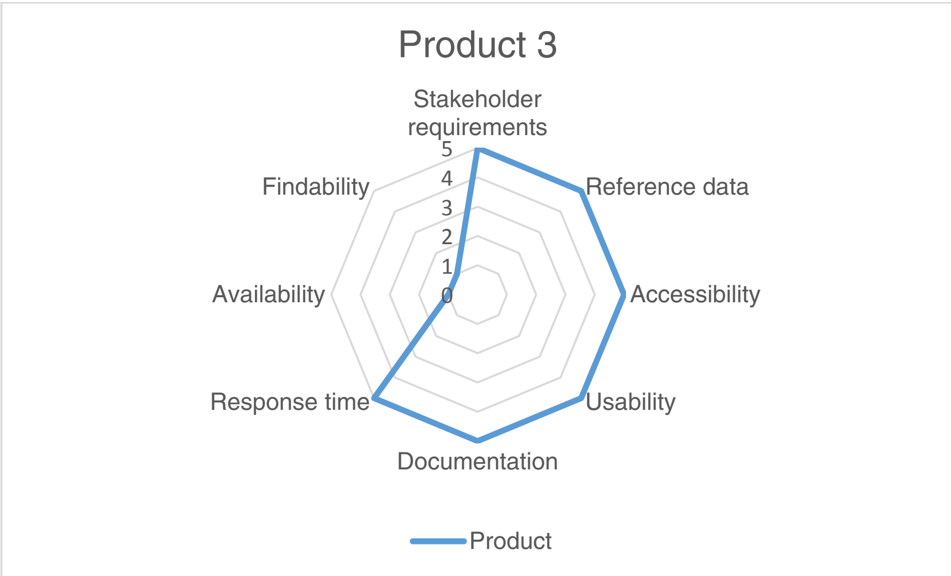


Figure XXVII: Scores sprint 5

Appendix VIII: Validation Techniques

After the execution of the queries for SL0 and SL1 level, the users per hour per day for the whole campus and the users per hour per day per faculty are known. However, there is no certainty that the amount of visitors calculated from the database is accurately representative of the real world. By validating the results, clear statements about the quality of the created information. Validation is a process which is very man-hour intensive, therefore it is not feasible to cower the whole campus area in the time allotted to the project. Instead validation is done at one or two faculties. It is necessary to find a method to determine the users per hour per day per faculty. This chapter discusses ways to collect this validation data.

There are different types of counting methods available, which can be divided in four groups:

- Systems using contact-type counters
- Systems using sensors
- Vision-based systems using cameras
- Manual counting (Senem, Ying-Li, & Arun, 2006, p. 1265)

I.XXVIII. Systems using contact-type counters

Figure XXVIII: A picture of the OV-gates of the NS.

Contact-type count systems are systems that block the path for a user. Examples of this type of system are public transport gates (OV-gates) (Figure XXXVII) or stadium access gates. The accuracy of contact-type count systems is high due to their barrier structure; users have to check-in to pass. The buildings at the TU Delft do not use a contact-type system, and installation costs are high. Therefore this solution cannot be used in the scope of this project to acquire the number of people entering and leaving.

I.XXIX. Systems using sensors

There are many different types of systems which use sensors to count people. An example of sensor system is infrared detection. A beam is sent to a reflector and bounced back to the sensor. The system counts a person when the beam is interrupted. By placing two sensors near each other the direction can be detected. Infrared systems are an inexpensive solution; the sensors are not expensive and they can be built from scratch. A potential problem might be to get permission to install the system in TU Delft buildings.

A shortcoming of an infrared system is that it can only function adequately is small entrances were only one person can enter or leave at a time. If the system is installed in a large entrance the consequence might be that two people passing b at the same time will be counted as one

I.XXX. Vision-based systems using cameras

I.XXX.I. Image processing: Facial recognition

Figure XXIX: Facial recognition techniques use spots to detect faces.

Another possibility is to monitor activity by taking pictures of building entrances. With the Microsoft Cognitive Services Face API persons can be identified as well as the position of the face in a picture (Figure XXXVIII). If multiple pictures are combined movement of the face can be tracked across the pictures. This information can then be used to decide if a person is entering or leaving a building. Taking pictures and uploading them to a third party (Microsoft, 2016a) without explicitly informing the populace is violation of the data protection directive. While it is possible to inform people of the activity, the uploading of the data to Microsoft servers cannot be excused.

I.XXX.II. Movement recognition

There are also possibilities to create a recording and track movement patterns from it. It is possible to send a video to a Microsoft service, use itrack cameras or to use new developments in the field (Senem et al., 2006), (itrack bv., 2015), (Microsoft, 2016b). Uploading the data to Microsoft is not allowed due to privacy issues. Buying a system from itrack is too expensive and by placing the system in a TU Delft Building the same problems will be encountered as with system using sensors. A final option is to create a custom system, however, the development of such a system would take more time than currently available for the project.

I.XXX.III. Thermal technology

Another approach is the use of thermal cameras. These systems are more reliable than systems based on the detection of movement patterns and have less privacy issues because people cannot be recognized (Traf-Sys, 2015). Therefore this method could be used to count the number of people. Buying a system is a too large financial investment as well as still needing permission from the University to install the system.

I.XXX.IV. 3D cameras

Using 3D Cameras can also be a solution, as they should be 98% accurate in detecting persons entering a building and can distinguish children from adults (itrack bv., 2015). However due to the technique being quite new, the systems are expensive as well as the software requiring heavy computing power.

I.XXXI. Manual counting

The fourth method to count visitors is manual counting. This technique consumes a lot of man-hours, but the technique is inexpensive in terms of financial costs. The accuracy of manual counting can be affected by human errors, though sufficient validation periods should reduce the effect.

I.XXXII. Conclusion

Uploading images and movies to a third party is not possible due to privacy issues. Buying counting systems will be a too large investment and will be difficult to install in TU Delft buildings. It is possible to make a system ourselves but this is not in the scope of the Synthesis project. Manual counting is quick to set up, requires little money and does not require difficult processing methods. Therefore manual counting is used to verify the query results. However FMRE installed itrack cameras in the TU Library allowing us to use some automatically generated data as well.

Appendix IX: Itrack validation

The main entrance of the TU Library is equipped with two itrack cameras, which provides an opportunity to count people entering or leaving the building. As the reliability of the itrack system is unknown, a validation is performed on two different days by manual counting. The goal is to determine the reliability of the itrack cameras in counting the total persons entering and leaving the TU Library.

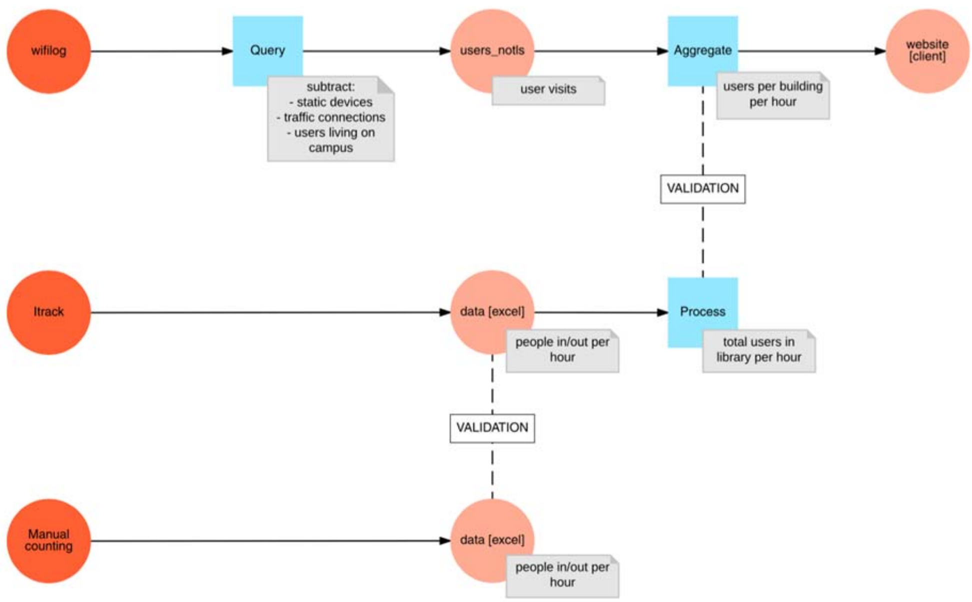


Figure XXX: Data validation workflow

The accuracy assessment can be summarized in three levels (Figure XXX). The final data from the queries which results in users per faculty per day, can be validated with the users entering and leaving the library. This results into an accuracy level. The level is affected by the quality of the itrack data. The data should be validated with a real world scenario in which the live counting of people is done.

I.XXXIII. Assessment Model

The quality of the itrack output is assessed according to Table X. As the itrack cameras log the movement patterns per hour, their validation should be performed for at least an hour. Since manual counting is easily prone to blunders, the validation should be repeated on a different day and executed by two team members.

	Day 11-May: Time 09:00- 10:00		Day 13-May: Time 09:00-10:00	
	Team member A	Team member B	Team member A	Team member B
Main Entrance IN	*Manual counting*	*Manual counting*	*Manual counting*	*Manual counting*
Main Entrance OUT				
Coffee bar IN				
Coffee bar OUT				
Sub Entrance IN				
Sub Entrance OUT				

Table X: Assessment model of the itrack validation.

I.XXXIV. Collect and analyse data

The results of the manual counting are shown in Table IX. Deviations in manual counting are likely to be human errors as a result from peaks in movements at certain moments during the session. The manual measurements are therefore averaged.

	11-May	13-May
Entrance IN	283	204
Entrance OUT	57	61
Coffee IN	137	83
Coffee OUT	136	72
Entrance 2 IN	339	239
Entrance 2 OUT	158	113

Table XI: Overview of the manual counting of the TU Library.

I.XXXV. Evaluate Results

The results from the manual validation are compared with the daily logs from the itrack cameras (Mantel, 2016). The results can be seen in Table XII.

	Day 11-May: Time 09.00-10.00	Day 13-May: Time 9.00-10.00
Main Entrance IN	283	204
Itrack Main Entrance IN	293	205
Deviation	*+3,5%*	*+0,5%*
Main Entrance OUT	57	61
Itrack Main Entrance OUT	109	75
Deviation	*+91%*	*+23%*

Table XII: itrack data validation at the TU Library.

I.XXXVI. Conclusion

The retrieved itrack data is considered viable for people entering the TU Delft Library, as a deviation below 5% is accepted. The data for people departing from the building is however deemed unreliable. A deviation of 23% could possibly be explained as validation errors, but a deviation of 91% cannot be justified. FMRE, as owner of the itrack scanners, is informed about this result. Currently, no explanation for this deviation is agreed upon, but a presumption has been made as incorrect scanner placement (Figure XXXI).

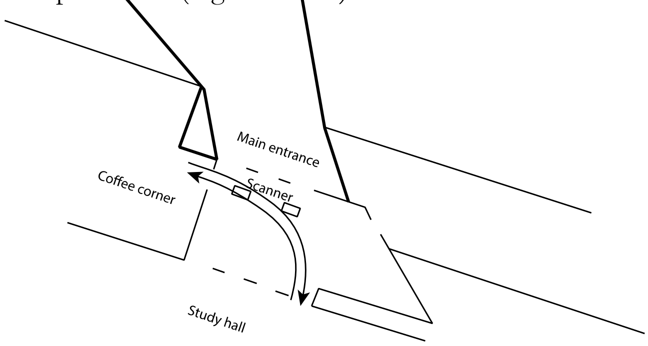
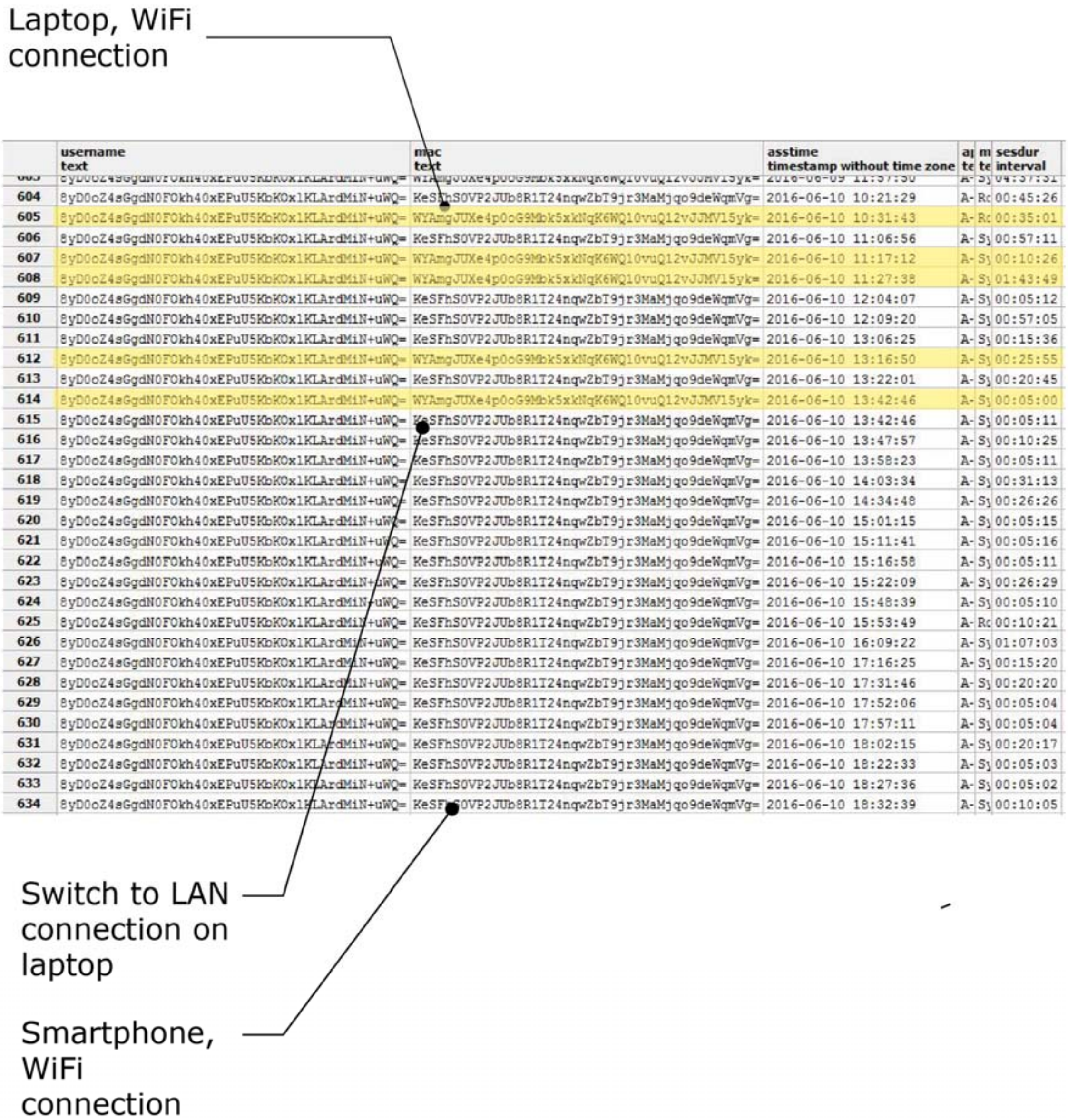


Figure XXXI: The presumption presented to FMRE is that the incorrect placement of the left itrack scanner misrepresents the actual people movement in and out of the TU Library, as people entering the coffee bar might also be detected by the system.

Appendix X: Logging LAN users

Users that are connected to the Eduroam network through a network cable are not recorded in the wifilog. The figure V shows a snippet of a field test done on the Faculty of Architecture. The field test was done by first identifying an individual group member (through historical data analysis). Then this user is followed by connecting to the Eduroam network through: 1) connecting via WiFi, and 2) connecting with a network cable. The wifilog shows the group member connected through WiFi highlighted in yellow. After the last highlighted row the WiFi connection was replaced by a LAN connection. There are no records in the wifilog table of the LAN connection showing that LAN connections are not recorded by the system.



	username text	mac text	asstime timestamp without time zone	a te	m te	sesdur interval
604	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 10:21:29	A-	Rc	00:45:26
605	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	WYAmgJUXe4p0oG9Mbk5xkNqK6WQ10vuQ12vJJMV15yk=	2016-06-10 10:31:43	A-	Rc	00:35:01
606	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 11:06:56	A-	Sy	00:57:11
607	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	WYAmgJUXe4p0oG9Mbk5xkNqK6WQ10vuQ12vJJMV15yk=	2016-06-10 11:17:12	A-	Sy	00:10:26
608	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	WYAmgJUXe4p0oG9Mbk5xkNqK6WQ10vuQ12vJJMV15yk=	2016-06-10 11:27:38	A-	Sy	01:43:49
609	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 12:04:07	A-	Sy	00:05:12
610	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 12:09:20	A-	Sy	00:57:05
611	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 13:06:25	A-	Sy	00:15:36
612	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	WYAmgJUXe4p0oG9Mbk5xkNqK6WQ10vuQ12vJJMV15yk=	2016-06-10 13:16:50	A-	Sy	00:25:55
613	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 13:22:01	A-	Sy	00:20:45
614	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	WYAmgJUXe4p0oG9Mbk5xkNqK6WQ10vuQ12vJJMV15yk=	2016-06-10 13:42:46	A-	Sy	00:05:00
615	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 13:42:46	A-	Sy	00:05:11
616	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 13:47:57	A-	Sy	00:10:25
617	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 13:58:23	A-	Sy	00:05:11
618	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 14:03:34	A-	Sy	00:31:13
619	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 14:34:48	A-	Sy	00:26:26
620	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 15:01:15	A-	Sy	00:05:15
621	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 15:11:41	A-	Sy	00:05:16
622	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 15:16:58	A-	Sy	00:05:11
623	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 15:22:09	A-	Sy	00:26:29
624	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 15:48:39	A-	Sy	00:05:10
625	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 15:53:49	A-	Rc	00:10:21
626	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 16:09:22	A-	Sy	01:07:03
627	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 17:16:25	A-	Sy	00:15:20
628	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 17:31:46	A-	Sy	00:20:20
629	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 17:52:06	A-	Sy	00:05:04
630	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 17:57:11	A-	Sy	00:05:04
631	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 18:02:15	A-	Sy	00:20:17
632	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 18:22:33	A-	Sy	00:05:03
633	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 18:27:36	A-	Sy	00:05:02
634	8yD0oZ4sGgdN0FOkh40xEPuU5KbKOx1KLArdMiN+uWQ=	KeSFhS0VP2JUb8R1T24nqwZbT9jr3MaMjqo9deWqmVg=	2016-06-10 18:32:39	A-	Sy	00:10:05

Figure XXXII: Field test on LAN logging shows that LAN users do not get recorded

References

Hillson, D., & Simon, P. (2007). *Practical Project Risk Management: The ATOM Methodology*: Management Concepts.

Hubbard, D. W. (2009). *The Failure of Risk Management*. New Jersey: John Wiley & Sons Inc.

Leijten. (2014). *Complex system structures and project management*. Retrieved from Delft:

Mantel, L. (2016, 13-05-2016). [Interview].

PMI Standards Committee, & Project Management Institute. (1996). *A Guide to the Project Management Body of Knowledge*: Project Management Institute.

van der Spek, S., & Verbee, E. (2016). *Monitoring Flows and Occupation Patterns in Wi-Fi* (1.0 ed.). Technische Universiteit Delft: Technische Universiteit Delft.

Vrancken. (2014, 24-04-2016) *Complex system structures and project management*. CME8000.