Determine activity based on the classified identity of users by using Wi-Fi monitoring

Technical Report - Geomatic Synthesis Project
TRACK-id

IJ.D.G. Groeneveld
R. Sulzer
E. Theocharous
M.S. Tryfona
O.T. Willems
Y. Xu

Delft University of Technology

Determine activity based on the classified identity of users by using Wi-Fi monitoring

Technical Report - Geomatics Synthesis Project - TRACK-id

by

IJ.D.G.Groeneveld
R.Sulzer
E.Theocharous
M.S.Tryfona
O.T.Willems
Y.Xu

in partial fulfillment of the requirements of the Geomatics Synthesis Project of the
Master of Science in Geomatics for the Built Environment
at the Delft University of Technology,
to be presented on Monday June 13, 2016 at 12:00 AM.

Project duration:	April 18, 2016 – June 17, 2016	
Project committee:	Drs. W. Quak,	TU Delft, coach
	Dr. ir. S. C. van der Spek,	TU Delft, project coordinator
	B. Valks M.Sc.,	TU Delft, FMRE

This report is in compliance with Data Protection Directive 95/46/EC, therefore certain parts might not be made public.

An electronic version of this report is available at http://repository.tudelft.nl/.

TU Delft

Preface

During the final quarter of the first year of the MSc program of Geomatics for the Built Environment of the Delft University of Technology (TU Delft), groups of students work intensively on the Geomatics Synthesis Project (GSP) for 8 weeks. In the current project, they apply and use all the skills that they learnt and the knowledge they gained in the first year, to a real world project with stakeholders, external influences and organisational challenges. With the GSP the following goals are aimed at:

1. A synopsis of the different disciplines that were taught in the first year of the MSc program of Geomatics for the Built Environment at the TU Delft comprising the following courses:
    - GEO1001: Sensing Technologies for the Built Environment
    - GEO1002: Geographical Information Systems (GIS) and Cartography
    - GEO3001: Python Programming for Geomatics
    - GEO1003: Positioning and Location Awareness
    - GEO1005: Spatial Decision Support for Planning and Crisis Management
    - GEO1006: Geo Database Management Systems
    - GEO1004: 3D Modelling of the Built Environment
    - GEO1008: Geo Datasets and Quality
    - GEO1009: Geo-information Organisation and Legislation
    - GEO1007: Geoweb Technology
2. Practice teamwork in small groups.
3. Experience the entire geo-information process from project definition, over measurement, data processing and analysis, to presentation, delivery and application.
4. Encourage collaboration between students, staff and professionals.
5. An innovative project extending the ‘state of the art’ with new fields of exploration.

The two topics of this year’s GSP are the tracking of pedestrians in Dordrecht and the Wi-Fi tracking of users on the campus of the TU Delft in all its aspects. From the 3 groups that work on the Wi-Fi tracking on the TU Delft: one group will look at the occupancy of education space, one group will look into the movement/trajectory of users and one group will look at activities/behavioural patterns.

This project looks into the activities of users and their behavioural patterns, mostly outside of the opening hours, and provides answers to the questions of the stakeholders. The GSP will be an intensive 8 week project divided into 3 phases: the first phase (baseline) where the team sets the organisational structure and identifies the requirements, the second phase (mid-term), where a conceptual analysis is prepared, in order to go into technical depth and explore the ways to get the answers of the stakeholder, and the third and final phase, where the actual analysis is performed to provide the answers to the questions of the stakeholder and a total and complete workflow on how to get the answers is presented during a seminar and the delivery of the final technical report.

IJ.D.G. Groeneveld, R. Sulzer, E. Theocharous, M.S. Tryfona, O.T. Willems & Y. Xu
Delft, June 2016

Acknowledgements

In performing our project, we required the help and guideline of some respected persons, who deserve our greatest gratitude. The completion of this project gives us much pleasure. We would like to show our recognition to Drs. Wilko Quak, our project supervisor of the Synthesis Project of the MSc Geomatics for the Built Environment of TU Delft, for giving us a good guideline for this project throughout numerous consultations.

In addition, a thank you to Ir. Edward Verbree, associate professor of the MSc Geomatics for the Built Environment of TU Delft, who provided us with very useful suggestions and remarks. The help of PhD candidate Lorenzo Dalla Corte of TU Delft's Faculty of Architecture and the Built Environment, at the privacy issues is really appreciated. Moreover, the advices and the support of Mr. Bart Valks, representative of Facility Management and Real Estate of TU Delft, for this project is acknowledged. We also thank the Information and Communication Technology (ICT) department of TU Delft for the dataset that it provided us. We would also like to thank Stefan van der Spek for the subject of our project and for the opportunity to expand our horizons on the Wi-Fi monitoring field.

Many people, especially our classmates and team members itself, have made valuable comment suggestions on this proposal, which gave us an inspiration to improve our project. We thank all the people for their help directly and indirectly to complete the TRACK-id project.

The authors

Executive Summary

The executive summary is written in the style of a paper in order to be submitted to a conference, and can be found on the next page.

DETERMINE ACTIVITY BASED ON THE CLASSIFIED IDENTITY OF USERS BY USING WI-FI MONITORING

IJ.D.G. Groeneveld, R. Sulzer, E. Theocharous, M.S. Tryfona, O.T. Willems, Y. Xu
Team Coach: Wilko Quak

KEY WORDS: Wi-Fi monitoring, Localisation, Markov model, Activity, Real Estate Management

ABSTRACT:
The Wi-Fi technologies are used in everyday life on numerous applications that detect the crowd information for commercial, security and other reasons. The Wi-Fi monitoring can be used for tracking people when they are moving along different access points. The results from the Wi-Fi monitoring can provide the location of the users in an area and therefore, useful information can be extracted. The goal of this project is to recognize the activity of different users for different sessions of a Wi-Fi network. The Wi-Fi dataset that is used, is acquired from the Wi-Fi network of the Delft University of Technology (TU Delft). Initially, the estimation of the users' occupation is determined with the use of a Markov model with the information that is derived from the Wi-Fi dataset. Their possible identity is used, in order to estimate the activity that a user is probably doing at a specific location of the research area. The results on the use of the research area, are calculated and visualised in different spatial levels, campus, building and floor level. The use of the building complex of the TU Delft Campus, is examined during irregular hours, to allow efficient real estate management and provide security solutions.

1. INTRODUCTION

The use of Wi-Fi networks by mobile devices increased rapidly over the last decade. Each mobile device has a unique Media Access Controller (MAC) address, which is emitted by each mobile device during its search for Wi-Fi access points. The MAC address is picked up at an access point, when the Wi-Fi enabled device is in range. This information can be used for tracking people when they are moving along different access points. The results from the Wi-Fi monitoring can provide the location of the users in an area and therefore, useful information can be extracted from such a venture.
The aim of this project is to recognize the activity of different users in an area through the Wi-Fi monitoring. First of all, the estimation of the user's occupation is calculated from the use of a Markov model with the information that derive from the Wi-Fi dataset. Their identity is used in order to estimate the activity that a user is probably doing.
The main question of this research is:
To what extend and how reliable is it possible to determine the activities of individuals through the users' characteristics that can be derived from a Wi-Fi network?
This project is focusing more on the use of the facilities of the research area during irregular hours. The irregular hours are specified as the hours outside of the opening hours of the buildings, not including exceptions like extended opening hours or events. The use of the buildings is examined during irregular hours, to allow efficient real estate management and provide security. The Information Communication Technology (ICT) department of Delft University of Technology (TU Delft) provided a database dump of the Wi-Fi network of TU Delft. These data are analysed, in order to get useful and valuable information about the usage of buildings and the activities in buildings.
The Wi-Fi tracking technology is analysed and the basic characteristics of the system that is used for this project, are presented in chapter 2 and 3 respectively. The methodology that is conducted in this project is described in chapter 4. Moreover, in chapter 5, some basic outcomes are analysed. Furthermore, some conclusions and recommendations for future work are specified in the chapters 6.

2. WI-FI TRACKING

Wi-Fi tracking uses wireless positioning techniques that refer to radio-navigation methods and rely on distance estimation (Mautz, 2012). Wi-Fi is a communication that uses radio waves over the air. It includes the end-user devices, such as smartphones, laptops etc., the radio frequency spectrum, and the infrastructure of the Wi-Fi network. The basic components of Wi-Fi service are depicted in Figure 1:

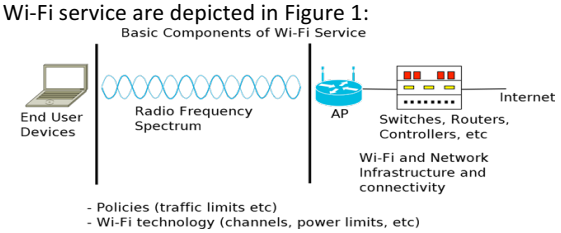


Figure 1: Basic components of Wi-Fi service (University of Waterloo Campus Community, 2013).

Wi-Fi tracking is not a new technology, but it is one that gathers momentum for use in the daily routine of people. For example, shop owners use it to get more insight into the usage patterns of customers and by passers, while free Wi-Fi hotspots are utilised to adjust advertising and services. Some existing localisation techniques are the RSSI and trilateration based, the fingerprinting based, the angle of arrival based and the time of flight based, with the first two rated as the most common used techniques.
The Wi-Fi tracking technologies have many benefits for companies and for society at large, but the quality of the results can be interfered by several reasons. Nowadays, there are applications that change the Wi-Fi MAC address of a device. This process is called MAC address spoofing (or MAC spoofing) and the users that use these kind of applications in a Wi-Fi network appear to have more devices than the usual

number of devices per person. Moreover, the dynamic environments and the different materials of the floors, the internal walls, the windows and the doors can cause reflection and fading to the signal. Interference on the transmission path of the Wi-Fi signal are also related to Radio Frequency and Electrical causes. When another device is using the same frequency in an area, even at low power levels, the Wi-Fi network in that area can stop working until the end of the interference. Devices that can provoke such problems are the cordless phones and the microwaves. The impact that electrical interference has on the signal, depends on the proximity of the electrical device to the wireless access point (Harwood, 2009). Electrical interference comes from devices such as computers, refrigerators, fans, lighting fixtures, or any other motorized devices. Concerning the complexity of the technology of Wi-Fi tracking and the several reasons of the interference of the signal, the accuracy of the results of Wi-Fi tracking needs to be distinguished through validation techniques.

3. DATASET LIMITATIONS

The dataset has been provided by a third party source, hence it is not possible to select the tracking technologies and methodologies. In this chapter, the infrastructure that the third party utilised to gather along with the validation, the accuracy and the representativeness of the tracking system, is discussed. Finally, an insight about the data protection that is relevant with the data and methodologies that are used, is presented.

3.1 Validity and accuracy

Due to the design of the tracking infrastructure, if a user is only passing by an access point and gets scanned, the user gets a minimum connection time of five minutes. Furthermore, the system is designed in a way that a device can only maintain a single connection at a time, thus, if a user moves inside the five-minute period of his last access, he will not be tracked to the new position.
Due to the nature of Wi-Fi tracking, a user that is detected in a certain access point, is not necessary located around that access point. For example, a user might be passing outside a building and he might be scanned by an access point inside it, when actually he is not there. Moreover, false location of scanned users can also happen between floors of a building.

3.2 Representativeness

The representativeness of the data collection reflects what categories of users can the implemented Wi-Fi tracking system identify, or what categories of users it cannot.
The first big category is the users that were not connected to the Wi-Fi network. It consists of smaller categories, a) users that use old technology, b) users that are not part of the academic network that has access to the Wi-Fi network, c) users who consciously turned off their Wi-Fi devices, d) users connected to the network through wired connection.
The second category refers to places that there are no available data in the given database. The missing data accounts for, either buildings that are excluded for security reason, e.g. Nuclear reactor, or for places that are not covered by access points, hence no information is collected.

3.3 System of access points

With the implemented system of access points, every device is usually detected in only one access point. Thus, the processes of fingerprinting and trilateration, that would allow the determination of the position of a device, cannot be carried out. Therefore, it is assumed that the detected devices are located to the specific access points that they are detected.

3.4 Data protection

According to the Data Protection Directive (DPD) (Directive 95/46/EC) 'Personal data is any information relating to an identified or identifiable natural person'. The data collected from the Wi-Fi network are considered personal data, even if the username and MAC address of each user and device is hashed, hence the data must be treated accordingly. Currently, DPD secures the data privacy of individuals against unlawful use. According to DPD, to process private data, one has to have valid ground to do so. In the current project, scientific purposes are sufficient reason, yet data must be processed with respect to the Principles mentioned in the DPD and with respect of the Data subject rights.

4. METHODOLOGY

The data structure that is maintained in the database system, is presented below. Furthermore, the pre-processing of the data, in order to be suitable for analysis and the steps of the analysis of this project, are described.

4.1 Data structure

The Wi-Fi monitoring data of this project is provided from the ICT department of TU Delft in a database system. On this system all the data is accessed, processed and analysed using different queries. The data structure as it exists within the database, it is shown in Figure 2 with each table that is created during the queries. Furthermore, these tables reflect the milestone products and steps in the pre-processing and analysis.

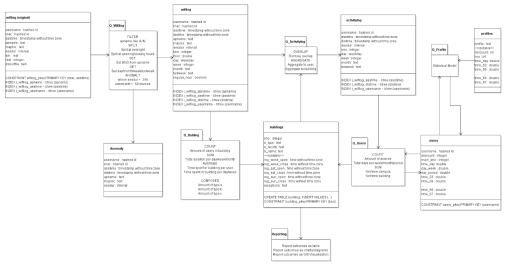


Figure 2: Class diagram with processing queries.

The starting dataset (wifilog table from Figure 2) of this project is the dataset of the Wi-Fi network of TU Delft, it is provided by the ICT department of TU Delft and contains all records straight from the source, as a dump of each day. Each record consists of a hashed user (username) that can be connected with different devices. Each device has its unique MAC address (mac) with an access time (asstime) to a specific access point (apname) for a certain duration (sesdur) and with the average signal strength (RSSI) and signal to noise ratio (SNR). Each access point has a location in a building on the campus (maploc). Finally, each record originates from a dump of a specific day (column importfile).

The columns RSSI, SNR and the column importfile have been removed from the starting dataset. Also, in this table the information that was implicitly contained in values, is extracted to explicit columns. For example, from the access point's column (apname), the building number (bno) is extracted, and from the access time (asstime) column, the columns weekday (i.e. Monday), month (i.e. April), week (i.e. week 32) and tudweek (i.e. week 3.10) are extracted. Moreover, the column regular_hour is added, which signifies whether a session was started and finished, while the corresponding building was closed. The above information is stored into a new table (wt3_wifilog), which is used during the pre-processing and the analysis of this project.
The rest of the tables that are shown in Figure 2, are created from the new table (wt3_wifilog) according to the needs of the pre-processing and the analysis steps that are mentioned below.

4.2 Pre-processing

Preprocessing is a procedure that takes place before the analysis of the data and it aims to render the data to be suitable for further processing and analysis. In this part, cleaning of data, data mining and filtering data take place. In Figure 3 the processes in pre-processing are arranged by the order of processing and dependencies.

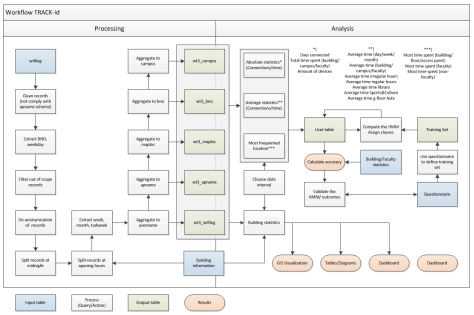


Figure 3: Workflow of queries arranged by order and dependency.

4.2.1 Cleaning of the data

The dataset that is provided, contains information which is not required for the processes designed, e.g. RSSI, SNR. All information not necessary is removed to reduce the data storage requirements.
In order to start the data mining and not make custom queries that have to take into account every deviation of the schema, records that don't comply with the schema, are removed. For instance, the access point name (apname) is consistent in most buildings, but a small number of access points have a value that does not follow the usual schema. All access point names not starting with a 'A- ', are being removed.
Some records have invalid session duration that results in consecutive sessions with an overlap period. Sometimes, a device is connected to more than one access point, which is impossible, based on the system of access points. More specifically, the first connection is still ongoing, while a second connection with the same MAC address is made, around five minutes later. Since it is not possible to tell which of the connections is correct, they are not removed. Instead, the records are flattened in different spatial aggregations.

4.2.2 Data mining

In this section, some data mining is processed, in order to get deeper into the problems and the solutions of the given Wi-Fi dataset. Information from the original dataset is extracted during this step, such as the building number (bno) that was encapsulated in the 'apname', and the weekday from the 'asstime'. These two columns are necessary to split the records of the dataset at opening hours.
A user in order to be connected with the Wi-Fi network, the connection can be done with Tunnelled Transport Layer Security (TTLS) or with Protected Extensible Authentication Protocol (PEAP). Occasionally, devices that use the TTLS authentication protocol can be connected as anonymous users during a connection to the Wi-Fi network. A large part of the device sessions of user anonymous can be de-anonymised and can have the not-anonymous username assigned to the device instead of anonymous. The sessions that remain (have no alternative non-anonymous user) are filtered out.
Next, the splitting of records is taking place, in order to improve the needed statistics. Records that span over more than one day are split up at midnight to separate the connections per day. The session duration is cut at midnight and the disconnection time is set to '24:00:00'. The rest of the session duration is added in a new record on the next day, or recursively split again if it passes another day. Splitting the records at midnight ensures that these columns always represent the true values of the session, when doing statistics about days/weeks/months.
Additionally, to the splitting at midnight hours, every record is checked for overlap with an opening or closing hour that is taken from the building table, according to the extracted building (BNO) of each session.

4.2.3 Aggregation

For profiling the user, in a later stage of the analysis, the distinction between devices is of no interest. The profiling of the users and not the profiling of the different devices of a user, is examined. Therefore, the different sessions of each MAC address/device of a user is aggregated. Different spatial aggregations are made on different spatial levels, campus, building, floor and access point level. An example of the aggregation of the different sessions is shown in the Figure 4.

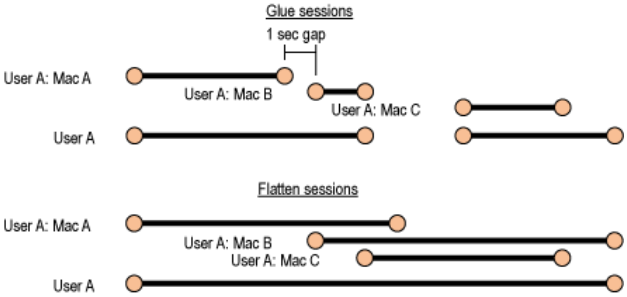


Figure 4: Visualisation of the aggregation process.

4.3 Analysis

After the pre-processing of the dataset, the analysis of the data is taken place. Into the steps of the analysis, a profiling and an activity model are utilized, for the framing of the users into the four different profiles and the assumptions of the activities for the different sessions of the dataset.

4.3.1 Profiling model

In order to distinguish the activity of a user, an occupation profile is assigned to each user. According to their occupation profile, their main building and the type of the building that they are located in, the activity of the specific session is determined. However, the main building of each user is determined as the building where that particular user spent his most time at.

The profile of the users is determined by using a Hidden Markov Model (HMM). A Markov model is a stochastic model that is used to model randomly changing systems, where it is assumed that future states depend only on the current state and not on the events that occurred before it. The stated property is characterised as 'memorylessness' or Markov property. Generally, this assumption enables reasoning and computation with the model that would otherwise be intractable. There are different Markov models used in different situations.

According to the Markov model, some training sets are defined considering the different profiles that need to be assigned to the users (student, academic staff, support staff or other). Those training sets are compared with the user's information that are derived from the dataset and the probability of each user to belong to each occupation profile is determined. Further, the user is assigned with an occupation randomly, based on the different probabilities (Petrushin, 2000 & Luhr et al., 2003 & Mühlenbrock et al., 2004 & Stamp, 2015). The process is illustrated in the Figure 5.

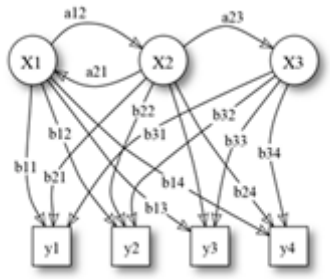


Figure 5: Visualisation of Markov Model (Wikipedia, Accessed: 9 June 2016).

4.3.2 Activity model

The activities of each individual, during irregular hours, is determined through a deterministic model, which takes into consideration the assigned user occupation, the user main faculty and the type of building the user is located each time. In the Table 1, the assumptions which connect all the above characteristics are being presented.

Table 1: Assumptions for activities.

	Build. type	Main build.	Study	Res.	Work	Other
Student	RES	ir.		x		
	FAC	ir.	x			
	EDU	ir.	x			
	SNC	ir.				x
	FMRE	ir.				x
Academic Staff	RES	ir.		x		
	FAC	yes		x		
		no			x	
Academic Staff	EDU	ir.	x			
	SNC	ir.				x
	FMRE	yes			x	
		no				x
Support Staff	RES	yes			x	
		no				x
	FAC	yes			x	
		no				x
	EDU	yes			x	
		no				x
	SNC	yes			x	
		no				x
	FMRE	yes			x	
		no			x	
Other Staff	RES	ir.		x		
	FAC	ir.	x			
	EDU	ir.	x			
	SNC	ir.				x
	FMRE	ir.				x

RES=research, FAC=faculty, EDU=educational, SNC=science, FMRE= facility management and real estate, ir.=irrelevant

5. OUTCOMES

In this chapter, the accuracy of the results and the outcomes of the project, are presented.

5.1 Quality control

In the following section the validation process, the overall accuracy and the methods to maintain accurate results are discussed.

5.1.1 Quality Control of classification

As seen in Table 1, during the preprocess there were 46.067 distinct users connected to Wi-Fi network, while, according to TU Delft reports there are 30.579 people, without taking guests into consideration. The 15.000 user difference is, as it is discussed, probably due to the large number of visitors that TU Delft has. Furthermore, over the total users recognized in the dataset there were 12% more students identified, 7% more academic staff and 4% less support staff, than in TU Delft statistical data. This leads to a RMSE of 5% (1528 users), which can be regarded as a very good result. However, it is not known if the users that are counted per occupation are the actual users of this group. Therefore, further accuracy measurements are given below.

Table 2: Profiling accuracy over total dataset users.

	Students	Academic Staff	Support Staff	Others	Sum (/others)
Calculated	23582	6815	3010	12660	33407
Reference	21545	6349	3128	unknown	30579
Difference	466	2602	-118	-	2950

Before the computations of the results of the Markov Model, the accuracy of the model is assessed. For the assessment of the accuracy 14 users and their occupations were collected and the model results are compared with the samples. In the tables 3 and 4, the results of the Markov model are illustrated, the students are recognized with 64% accuracy while all other classes are at 0%, for an overall 50% accuracy. The students were more closely simulated, because the training sets for students were based on actual students and not just

assumption, thus if the training sets improve, then the results should be better.

Table 3: Profiling results in known users.

ID	Real Occupation	Assigned Occupation
User 1	Student	Student
User 2	Student	Student
User 3	Student	Student
User 4	Student	Student
User 5	Student	Academic Staff
User 6	Student	Academic Staff
User 7	Student	Student
User 8	Support Staff	Student
User 9	Support Staff	Academic Staff
User 10	Academic Staff	Student
User 11	Student	Student
User 12	Student	Academic Staff
User 13	Student	Student
User 14	Student	Student

Table 4: Confusion matrix of classification.

		Ground truth data				
		Stud.	Aca. Staff	Sup. Staff	Other	Class. Overall
Classified Results	Stud.	7	1	0	1	9
	Aca. Staff	3	0	1	0	4
	Sup. Staff	0	0	0	0	0
	Other	1	0	0	0	1
	Truth Overall	11	1	1	1	14
	User Accuracy	63.63%	0%	0%	0%	

5.1.2 Accuracy of main building

For determining the activity of each user, its main building, its occupation and its scanned location are taken into consideration. To make an accurate statement, independent of the occupation classification of a user, the total numbers of faculty staff and students are compared to the numbers of the dataset. The guests of the university have to be filtered out, since official statistics do not contain guests of the university. Users that have less than four connection days, are classified as guests, leading to 13200 users being filtered out. After this procedure, the main building of a user is determined, as his main faculty, but only if he spent more than one hour there. The following table gives an overview of the accuracy of the home faculty or workplace determination.

Table 5: Accuracy of the determination of the number of users per building.

	3mE	Arch	CEM	EEMCS	IDE
Calculated data	4749	4435	4903	3693	3230
Reference data	5346	3698	4610	3529	2435
Difference	-597	737	293	164	795
	AE	TPM	AS	Other	Total
Calculated data	2743	1764	3613	1905	31035
Reference data	3163	1877	3840	2099	30597
Difference	-420	-113	-227	-194	438

5.2 Results

The results of the use of the TU Delft campus are visualised on different spatial levels and on different representations. The spatial levels that are used, are related to the campus, building and floor ("maploc") levels. Tables and graphs, a dynamic visualisation and a GIS and Web application are created during this project. In this section, some representative samples of the graphs on the three different spatial levels, are presented.

The Figure 6 shows the number of users during the regular and irregular hours at the TU Delft Campus from the 1st of April until the 31st of May. The largest groups of users of the TU Delft Campus in descending order are: students, academic staff, support staff, and others.

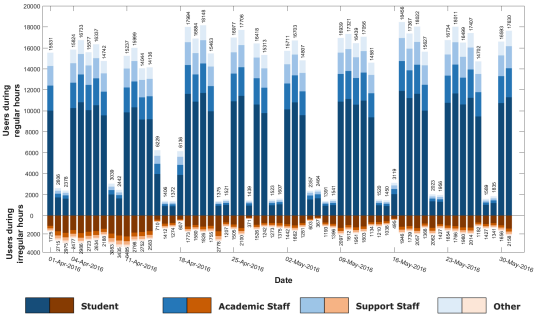


Figure 6: Number of users during regular/irregular hours at TU Delft Campus.

The Figure 7 shows the number of hours spent during the regular and irregular hours at the TU Delft Campus from the 1st of April until the 31st of May. The activities in descending order in terms of hours spent, are: study, work, other, and the least hours were spent at research.

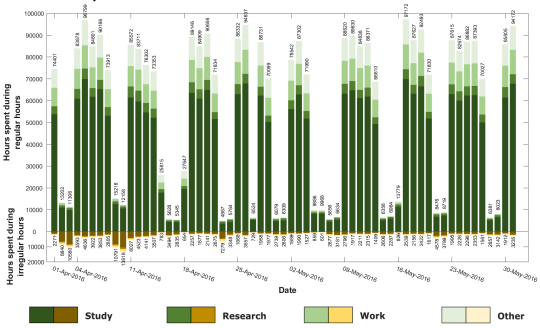


Figure 7: Hours spent during regular/irregular hours at TU Delft Campus.

As can be seen in Figure 6, the amount of users during the regular hours in the weekends is much smaller compared to the amount of users during regular hours of the weekdays. This is also reflected in the amount of hours spent, Figure 7. Regarding the irregular hours, the hours spent are usually higher during the weekends, than in the weekdays. Moreover, the differences in the number of users, between weekends and the weekdays, are little. Finally, during the irregular hours, the hours spent in the campus are varying a lot throughout the two-month period.

Depending on the various Dutch holidays and activities that took place in TU Delft Campus during the research period, it is possible to distinguish variations in the graphs, e.g. the low number of users at the TU Delft Campus on the 16th of May, which is a national Dutch holiday and most of the buildings of the campus are actually closed.

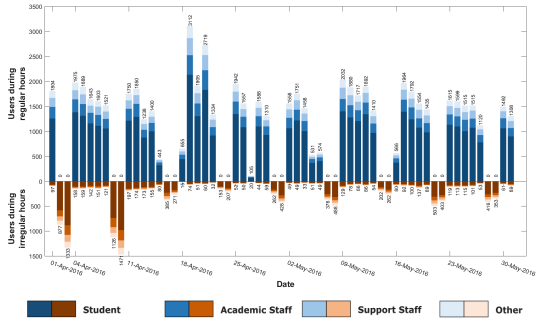


Figure 8: Number of users during regular/irregular hours at Aula conference centre.

As it is depicted from Figure 8, the number of users in Aula conference centre, during the irregular hours, on weekends is much larger than on weekdays. On Fridays there are always less people using the Aula conference centre than on the other normal weekdays.

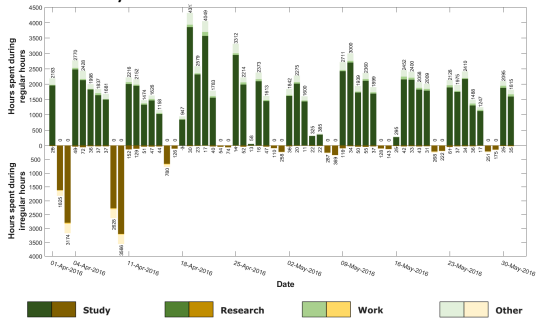


Figure 9: Hours spent during regular/irregular hours at Aula conference centre.

From Figure 9 it can be observed that during weekends there are users only during the irregular hours. Moreover, there are many users during the first two weekends, which were the exams period on TU Delft and students were studying. These are results of the exceptions that were not taken into consideration.

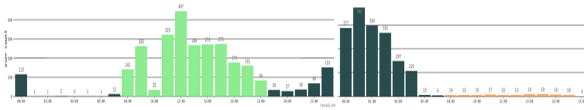


Figure 10: Hourly user count of Faculty of Architecture and the Built Environment on 22/04/2016 on the ground floor.

The example shown in Figure 10 represents the hourly user count on the ground floor of Faculty of Architecture and Built Environment. It can be deduced that an event started at midnight of Friday on the 22nd of April 2016 and ended at about 5am the next day (Faculty Party). The maximum number of the participants is estimated around 800. Moreover, about 500 people left between 4am and 5am (end of music show). On Saturday morning, on the 23rd of April 2016, there were about 15 people still at the same map location; these people are presumed as staff or students, who were probably responsible for cleaning the space of the event.

6. CONCLUSION

The use of Wi-Fi networks by mobile devices increased rapidly over the last decade. The MAC address of each device is picked up at an access point, when the Wi-Fi enabled device is in range. This information can be used for tracking people when they are moving along different access points. The aim of this project is to recognize the activity of different users in an area through the Wi-Fi monitoring. The general findings of the project are presented below. According to Table 5, the overall accuracy of the determination of the number of users per building is 94%. Moreover, through the Wi-Fi network, there were identified 438 distinct users more than the real users, excluding the guests, which is a deviation of 1.4% of TU Delft user statistics. Regarding the determination of the users' occupation using the Markov model, from Table 4, the accuracy of the process is 50%. Considering the above results, the pre-process and the analyses conducted to detect distinct users in a complex of buildings, are regarded of good quality and can be used further. On the other hand, the determination of user occupation is not accurate enough and further research is required. Additionally, the identification of specific events and exceptions on the opening hour of buildings can be identified by detecting irregularities of user connections. Finally, it is clear that through Wi-Fi tracking it is possible to extract information that will allow efficient real estate management and provide security solutions.

ACKNOWLEDGEMENTS

In performing this project, the help and guideline of some respected persons, who deserve our greatest gratitude, were required. The help of PhD candidate Lorenzo Dalla Corte of TU Delft's Faculty of Architecture and the Built Environment, at the privacy issues is really appreciated. Moreover, the advices and the support of Mr. Bart Valks, representative of Facility Management and Real Estate of TU Delft, for this project is acknowledged. The Information and Communication Technology (ICT) department of TU Delft is appreciated for the dataset that it provided for this project.

REFERENCES

DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 1995, Official Journal L 281, 23/11/1995 P. 0031 - 0050

Harwood M (2009) CompTIA Network+ N10-004 Exam Cram. 3rd Edition, Published Mar 13, 2009, Pearson IT Certification, Part of the Exam Cram series

Luhr S, Bui HH, Venkatesh S, West GAW (2003) Recognition of Human Activity through Hierarchical Stochastic Learning

Markov Chain. In Wikipedia. Retrieved 9 June 2016, from https://en.wikipedia.org/wiki/Markov_chain

Mautz R (2012) Indoor Positioning Technologies. ETH Zurich, Department of Civil, Environmental and Geomatic Engineering, Institute of Geodesy and Photogrammetry

Mühlenbrock M, Brdiczka O, Snowdon D, Meunier JL (2004) Learning to Detect User Activity and Availability from a Variety of Sensor Data, Xerox Research Centre Europe, Meylan, France

Petrushin V A (2000) Hidden Markov Models: Fundamentals and Applications Part 1: Markov Chains and Mixture Models

Petrushin V A (2000) Hidden Markov Models: Fundamentals and Applications Part 2: Discrete and Continuous Hidden Markov Models

Stamp M (2015) A Revealing Introduction to Hidden Markov Models

University of Waterloo Campus Community (2013) Factors affecting Wi-Fi service

Contents

List of Figures

List of Tables

List of Abbreviations

AP	Access Point
BNO	Building Number
CBP	College Bescherming Persoonsgegevens
DPD	Data Protection Directive
EU	European Union
TU Delft	Delft University of Technology
Eduroam	Education roaming
FMRE	Facility Management and Real Estate
GDPR	General Data Protection Regulation
GIS	Geographical Information System
GSP	Geomatics Synthesis Project
HMM	Hidden Markov Model
ICT	Information Communication Technology
LAN	Local Area Network
MAC	Media Access Controller
PEAP	Protected Extensible Authentication Protocol
RFID	Radio-Frequency Identification
RSSI	Received Signal Strength Indication
WLAN	Wireless Local Area Network
WPS	Wi-Fi based Positioning System
RMSE	Root Mean Square Error

1 Introduction

The use of wireless networks by mobile devices increased rapidly over the last decade. The amount of Wi-Fi Access Points (APs) soared, which make it possible for people to make use of Wi-Fi networks with their mobile devices at more and more places. Each mobile device has a unique Media Access Controller (MAC) address, which is emitted by each mobile device during its search for Wi-Fi APs. The MAC address is picked up at an AP when the Wi-Fi enabled device is in range. This information can for example be used for tracking people when they are moving along different APs.

The scope of this research is the determination of the amount of users and their activities outside of opening hours, in the Delft University of Technology (TU Delft) buildings. The Wi-Fi tracking technology is analysed and the scope of the research is presented in combination with the characteristics of the TU Delft campus and the main problem that this project is exploring to solve, in order to comply with the goals of the campus vision and the development of the TU Delft. The restrictions, the external limitations, the resource allocation, the performance prediction, and the technical risk assessment of this project, are described below.

1.1. Introduction into Wi-Fi Tracking

In order to have a better understanding of society and its behaviour within the (built) environment, policy makers and the general populace are interested in knowing where people are and what they do. There are some technologies that enable the tracking of people, by storing information that is send out to the connected network. Examples of such technologies are: the mobile cellular network, Bluetooth and Wireless Local Area Networks (WLAN) or Wi-Fi.

Wi-Fi is a communication that uses radio waves over the air. It includes the end-user devices (such as smartphones, laptops, etc.), the radio frequency spectrum, and the infrastructure of the Wi-Fi network. Connectivity to the Internet is usually a characteristic of a Wi-Fi network, but is not necessarily required (a closed local network for example). The basic components of Wi-Fi service are depicted in Figure 1.1.

Wi-Fi tracking is not a new technology, but it is one that gathers momentum for use in the daily routine of people. For example, shop owners use it to get more insight into the usage patterns of customers and passers-by, while free Wi-Fi hotspots are utilised to adjust advertising and services to appeal to a target audience. Wi-Fi tracking uses wireless positioning techniques that refer to radio-navigation methods and rely on distance estimation (Mautz, 2012). Some existing localisation techniques are the Received Signal Strength Indication (RSSI) and trilateration based, the fingerprinting based, the angle of arrival based, and the time of flight based, with the first two rated as the most common used techniques. The first is related to the measuring of signal strength from a device to several different APs and the

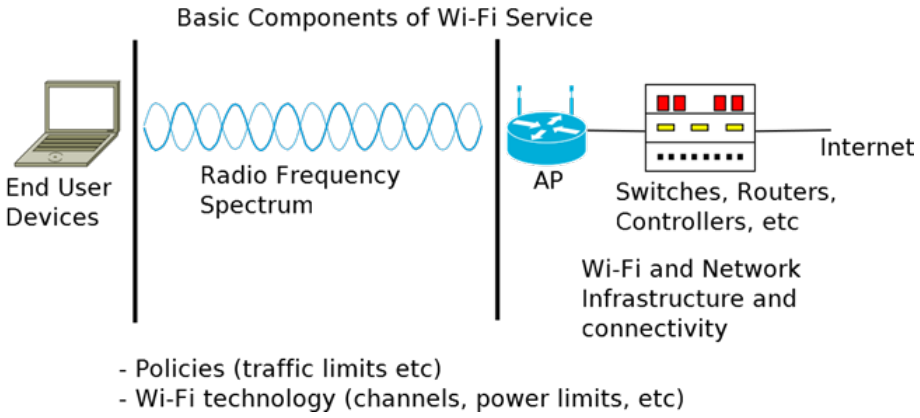


Figure 1.1: Basic components of Wi-Fi service (University of Waterloo Campus Community, 2013)

combination of that information with a propagation model to determine the distance between the device and the APs. Trilateration techniques can be used to calculate the estimated device position relative to the known position of at least three APs. RSSI measurements tend to fluctuate due to changes in the environment or the multipath fading. The fingerprinting technique is also RSSI-based, but it simply relies on the recording of the signal strength from several APs in range. This information is stored in a database along with the known coordinates of the device in an offline phase. The main disadvantage of fingerprinting is that in a dynamic environment the ”fingerprint” that corresponds to each location may change and an update of the new fingerprint in the database is needed (Mautz, 2012). The location of a device is linked with the location of a user and the location of a user can be related to many different factors that correspond to his behaviour, his personality etc.

Wi-Fi tracking has many benefits for companies and for society at large, but also a lot of pitfalls: big brother like applications, or ill practises. The Dutch ‘College Bescherming Persoonsgegevens (CBP, institute for the protection of personal information) went to great lengths to look into the possibilities and impossibilities of tracking and recently announced that Wi-Fi tracking in shops is no longer allowed, since it is an infraction upon the privacy legislation of The Netherlands. Therefore, it is no longer allowed for shop owners to collect uniquely identifiable MAC addresses, on the one hand because it involves personal data and on the other hand, because retailers don’t have the proper knowledge to safely store the information using encryption keys and other security measures (CBP, 2015), (Verlaan, 2015).

1.2. Interference problems of Wi-Fi tracking

There are applications that change the Wi-Fi MAC address of a device, this process is called MAC address spoofing (or MAC spoofing). For tracking technologies that just capture and store MAC addresses this will result in untrackable devices. However, if a users has to log in to a network or applications it can be detected even with using MAC spoofing, because in the dataset, the device corresponds to a username with more than the expected amount of devices (around four devices). These kind of cases are resolved in this project with the aggregation of the session durations of the different devices of a user (Section 3.2.3).

The calculation of the number of users in a building, on different floor levels (maploc), have some errors. These errors are related to the materials of the floors, the internal walls, the windows and the doors, and also on obstructions in the environment (furniture, people). The different materials and the dynamic environments can cause reflection and fading of the signal. Therefore, a person might be detected at an AP in a building that is on a different location, than the closest (expected) AP. In this case, the accuracy of the number of users that are detected on floor level need to be distinguished with validation techniques (Section 4.2).

More factors that can cause interference on the transmission path of the Wi-Fi signal, and therefore

introduce errors in the results of this project, are related to the Radio Frequency and Electrical disturbances. The first is linked with the Radio Frequency range of the 2.4GHz band that the wireless technologies use. This range can be the same as the range of the Wi-Fi scanners and therefore might be recorded in the dataset. Devices that can provoke such problems are the cordless phones and microwaves. When another device is using the same frequency in an area, even at low power levels, the Wi-Fi network in that area can stop working until the end of the interference. The second is electrical interference that comes from devices such as computers, refrigerators, fans, lighting fixtures, or any other motorised device. The impact that electrical interference has on the signal depends on the proximity of the electrical device to the wireless AP (Harwood, 2009).

As more research is done into the differences of chipsets, the Wi-Fi connectivity of different device brands (Apple, HTC, Samsung, etc.), and operating software (Android, iOS, Windows), a better understanding can be formed on how a device's signal strength influence connection and how connections are established. With known deviations from such devices, it might be possible to detect and improve the quality of the Wi-Fi tracking results.

1.3. Rhythm of the campus project

What is the Wi-Fi tracking project 'Rhythm of the Campus' of TU Delft all about? Education roaming (Eduroam) has been around for a while. It is the Wi-Fi network that all students, staff and researchers can connect to, in order to work in their universities and have access to other universities networks with the login provided by their home organisation. TU Delft is one of the participating universities and provides a connection service to eduroam with 1700 APs spread out over 30 buildings. More than 40000 unique users (both human users and static non-human users), generate almost 150000 connections on average per day. Keeping track of how long users are connected to which AP with which device brings great challenges, but can also bring great opportunities. With the sheer size of the network, the Wi-Fi tracking data can be considered as big data. This data was previously, until now, not stored. Since March 2016, the TU Delft decided to store this data. In order to make the data harder or even impossible to trace back to a single user, it is hashed and made sufficiently anonymous. Since the tracking started, it will continue to do so for an indefinite period of time. After a certain period, it can be evaluated to maybe become open data, if the tracking data allows it. That is, if there is no conflict on the matters of privacy or any other legal restrictions. The Geomatics Synthesis Project (GSP) will shed light on these delicate topics. Furthermore, the task is set to get a grip on the big data, with all its errors and inconveniences and extract useful information for stakeholders and researchers alike (Abedi et al., 2015).

Every day the Information Communication Technology (ICT) department of the TU Delft will provide the GSP groups with a database dump of the eduroam sessions/connections. Then, this data is analysed in order to get useful and valuable information about the usage of buildings and the activities within buildings.

1.4. Background to the problem

The aim of this project is to provide the Facility Management and Real Estate (FMRE) department of the TU Delft with information about the use of the facilities on the campus during irregular hours. The irregular hours are specified as the hours outside of the opening hours of the buildings, not including exceptions like extended opening hours or events. The information that is derived from this project is of key importance to FMRE, in order to provide recommendations and policy support concerning the use of real estate and the security of the campus during these irregular hours.

1.5. Campus vision and development of TU Delft

The TU Delft scores high in the international university rankings. This leading position demands buildings and facilities of high quality. Therefore, in the coming years, the TU Delft will invest heavily in the campus and its property in order to guarantee this quality also in the future (TU Delft, 2013a).

The property of the TU Delft must support excellent education and research. When property becomes technically and functionally outdated and does not keep up the pace with the improvements in performance, this can be a threat to the continuity of parts of the TU Delft and therefore to its leading position. The TU Delft would prefer to spend its money on the primary tasks of education, research and valorisation. Therefore, the TU Delft is utilising the available floor area as efficiently as it can, thereby increasing the multifunctional use of buildings and building as sustainable as possible (TU Delft, 2013a).

At the moment each faculty has its own teaching rooms, study places, research areas, offices, refreshment areas, and catering services. It is possible to use space more efficiently and thereby reduce costs, if the TU Delft does not necessarily need to maintain all this different functions in each faculty (TU Delft, 2013a). This leads to the flexible use of space, which is not only beneficial in terms of efficiency, but it is also making the campus more alive.

FMRE is responsible for the management of the property of the TU Delft and has to implement the campus vision that is mentioned in section 1.3. In order to make decisions, FMRE wants to know how many people at what times are using a certain building and what these users do. For the determination of the activities of the users after the closing hours, the occupation of users must be known in order to assume their activities, i.e. students that studies, an academic staff member that does laboratory work or a support staff member that works. This information can then be used to adjust policy on buildings, and for example, to adjust the opening hours or reassign security persons.

1.6. Scope of research

The purpose of this project is mainly to answer the questions of our stakeholder FMRE. Their main interest lies into what extent the campus is used during the irregular hours. Therefore, the scope of the research of this project is the determination of the amount of users and their activities outside of opening hours, in the TU Delft buildings marked by FMRE. Initially, the numbers of campus users, during the irregular hours, are determined. The number of the different users is data mined from the dataset, the results are visualised for different periods of the dataset, and are compared between the regular and irregular hours. Furthermore, the duration and the activities of the different users are explored. The information about the exact location of the APs within the buildings is not available. Thus, the identification of a user's activity is not possible based on the dataset, since activities are not directly measured, only the stays of users is. With only the information that is given about a user, namely the access time and duration of a connection session, and the information of a sessions location (on building level), it is not possible to define activities just yet. However, the aggregation of the multiple connections per user, allow the determination of a user's behaviour, from which a user profile can be extracted. After that, the occupation of each user is classified and his/her activity is estimated based on the different occupations. Therefore, with a user's typical behaviour and occupation, his general location, and the access time of a session, the activities, during the irregular hours, can be determined, this angle on activity through identity is used to estimate what a user is actually doing (Christensen et al., 2014).

Additionally, in this project, the general accuracy and representativeness of the TU Delft eduroam network, as well as the infrastructure of the AP system, are investigated. By dealing with the privacy of the data, a general conclusion on the viability of publishing the data with personal concerns taken into account. The MoSCoW chart (see Table 1.1) ranks the different options of our research.

The general methodology of preprocessing, analysis, quality control and presentation can be used in different Geographical Information System (GIS) projects. The methodology applied in this project can

Table 1.1: MoSCoW diagrams of research

MUST	SHOULD
How many users were in each building during irregular hours? How long did (each) user(s) stay in the building during irregular hours? Create a profiling framework. Report outcomes as tables. Report outcomes as diagrams. Validate assumptions.	Who did stay in the building during irregular hours? Where were users in a building during irregular hours? What were users doing in a building during irregular hours? How representative is the Wi-Fi data? Is the Wi-Fi data, personal data? Report outcomes as GIS visualisation.
COULD	**WON'T**
Automate the process to create outcomes with manual input of client (tables, charts/diagrams, GIS visualisation) Assess accuracy of location determination?	Create a dashboard/application to get outcomes. Assess quality of AP system

be used in a broader field. This project's methodology follows four main steps that can be used in different projects. The four steps are related to the cleaning of the data to make it fit for purpose, the counting of people that are connected to the network in a certain area, the framing of people based on characteristics that can data mined from the dataset, and identify possible activities within buildings, based on the users that appear there. Finally, a dashboard is provided, where parts of the dataset can be picked and the corresponding results are visualised in graphs. Projects that are related to different wireless technologies, such as Wi-Fi-based Positioning System (WPS), Bluetooth and Radio-Frequency Identification (RFID) can use the specific approach. The framing methodology, namely the Markov model (sections 2.3, 3.3.1 & 3.3.2), is a method that is used in a wide range of fields such as physics, chemistry, medicine, music, game theory, sports and also in activity recognition and pattern recognition. The recommendations in chapter 8 are on the one hand about the improvement of the applied methodology of this research in case the methodology is used on other projects about wireless (tracking) technologies. And on the other hand about tracking issues, the validity and accuracy of the results, the representativeness of the dataset and the suitability of the system of APs.

1.7. Restrictions and external limitations

The Wi-Fi data is gathered by a system of APs spread out over the whole campus. The system is installed by a third party. The ICT department provides a log of every connection that is successfully made with the eduroam network. A successful connection can only be made by a user that has a login from TU Delft or any other participating university of eduroam. The data only includes connections that are made via Wi-Fi and not with a wired Local Area Network (LAN) connection. To what extent the given dataset represents all the people from TU Delft, is evaluated in the validation section (section 4.1).

A connection is saved as a login of a user, with a certain device at a certain AP. Additionally the access time and session duration are saved. According to the dataset, the sensing interval is considered to be 5 minutes. After a five to six minute interval, each AP checks if the connection is still ongoing. If so, the session time keeps counting. If not, the connection is closed, with a session duration that ends at the end of the five to six minute interval, even if the session was closed before. This means that the precision of the session duration is around five minutes.

Moreover, the dump of the data is provided every day, which means that sessions that overlap with the time that the dump is made, can contain errors. Since the team does not have influence on the data collection, some false records have to be and are removed (details are explained in chapter 3).

On average there are around 60 APs in each building of the university, the Wi-Fi network should therefore provide coverage in every room, in line with the scope of this research. This does, however, not mean that the location of a user can be pinned down accurately to room level. Since a user can only be connected to one AP at a time, it is not possible to calculate the accurate location of the user. The

distance to an AP can only roughly be estimated by the signal strength provided in the dataset, which is an average over a whole session. Without further information on the coverage of an AP, the location of a user can only be determined on building level or on aggregated areas inside a building e.g. floor. Due to the unavailable AP location information, the the spatial resolution is limited to floor level as room level is not available.

The access time and the location information from the dataset might still be personal data, even if they are aggregated on building level. In order to protect the privacy information of the users, the personal username and the MAC addresses are encrypted. Further discussions about data protection follow in chapter 5.

1.8. Resource allocation and performance prediction

The resources of the project are divided into two parts, the database dump that is provided on a PostgreSQL server with all the characteristics described above and the people that are involved in the project. In the ideal case, the server should be updated with new data daily. The research is done by 6 research members that work a total amount of 1680 hours over a period of 8 weeks (280 hours per person). With the limited amount of resources, the application outcome will be a minimum viable product. A detailed plan of the following weeks is given in a GANTT chart (Appendix E).

1.9. Technical assessment risk

The risks that can endanger a successful result of this project are assessed. Table 1.3 shows the effect, cause, likelihood, severity and the consequential importance of possible risks with a classification of the different levels of likelihood and severity as is given in Table 1.2.

Table 1.2: Risk classification of risk map

Likelihood scale	Severity scale
1 - This cause is unlikely to happen.	1 - The impact on the project is very minor.
2 - This cause could conceivably happen.	2 - The impact on the project is noticeable.
3 - This cause is very likely to happen.	3 - The impact on the project is severe.

Table 1.3: Risk map of GSP

ID	Risk Item	Effect	Cause	Likelihood (L)	Severity (S)	Importance	Action to minimize risk	Owner
	Describe the risk briefly	What is the effect on any or all of the project deliver-ables if the cause actually happens?	What are the possible cause(s) of this risk?			L*S	What action(s) will you take (and by when) to prevent, reduce the impact of, or transfer the risk of this occurring?	Who is responsible for following through on mitigation?
1	Missing data.	Problem in the Hidden Markov Model (HMM).	Problems in the server.	1	1	1	Validate the data.	ICT
2	Wrong in-terpretation of the way the system works.	Invalid statistics.	Not enough research done.	2	2	4	Validate the data.	Everyone
3	Wrong in-terpretation of data.	Project failure.	Not enough research done.	1	3	3	Data mining and research.	Everyone
4	Flaws in the assump-tions.	Poor accuracy.	Wrong decisions.	2	3	6	Verify the results.	Everyone
5	Data structure is not feasible.	Loss of time.	Poor planning.	2	2	4	Better or-ganisation.	Everyone
6	Run out of time.	Poor project results.	Misjudge-ment of time.	2	3	6	Follow the project guide's deadlines.	Project Lead

1.10. Structure Overview

The following chapters give a more detailed description of the approach of the team and the choices that were made. The steps that the researchers have made, to get from big data to useful information, are explained in detail. First of all, the general methodology is described in chapter 2, followed by a detailed explanation of the step-by-step applied methodology in chapter 3, starting with the data structure, where the purpose and the structure of all tables is explained. Then secondly, the preprocessing steps are explained in detail, where insight is given into the cleaning of the data and the preliminary data mining of the dataset. Then, light is shed on the approaches that are applied during the analysis in order to get additional information about users and buildings, which is then used for the user's profiling and finally identifying activities. In addition the this, the validity and accuracy of the given dataset, the representativeness of the results of the project, and the system of APs are assessed in chapter 4. The data protection that corresponds to the privacy issues, is presented in chapter 5. Problems and solutions during the process of this research are described in chapter 6. Conclusions and recommendations are described in chapter 7 and 8 respectively. And last but not least, the outcomes of the research for FMRE are told in the story of the data in appendix A.

2

General Methodology

In order to understand worldly and (built) environment phenomena, the phenomenon has to be captured (acquired), processed, analysed and presented into useful information. All the steps in the process should be done with a certain quality control. These steps are cyclic and are part of the process of geo-managing (Lemmens, 1990). These consecutive steps are shown in Figure 2.1.

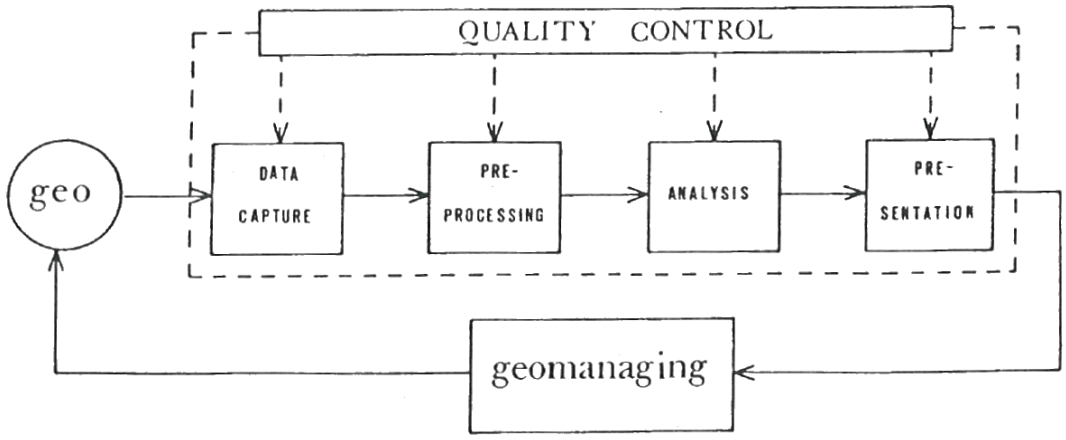


Figure 2.1: Steps in the production and use of geo-data (Lemmens, 1990)

For Wi-Fi tracking these general steps are applicable and have their specific approach/application for tracking, which is analysed in the following sections.

2.1. Data Capture

During the data capture step, the phenomenon is acquired by sensors. In the case of the 'Rhythm of the Campus' project, this step is done by a 3rd party. The individual users on the campus are captured and their sessions are stored in a database (a session of a user's device with a start time at a certain AP).

2.2. Pre-processing

Before the analysis of the data, the data needs to be clean and valid and fit for use/purpose. Therefore, the data needs to be processed. It is up to the researchers how to clean records that are considered invalid and interfere with the analysis methods. For Wi-Fi tracking the data should be synchronised, don't contain false records (i.e. double records and multiple connections to different APs if the system does not support this), and be useful for analysis without much manual editing.

The next step within processing is the extraction of implicit information into explicit columns (for easy access), this can be done by rewriting complex and combined field into separate fields, or use other datasets and fuse them on the table based on a common field or shared association. For example, implicit location information, as part of an AP to an explicit location description.

The last step in processing includes the creation of a subset of the original dataset by filtering irrelevant records. The dataset can be yearly, monthly, weekly, daily etc. depending on the needs of the analysis, visualisations ,and the comparison of the different results. Without throwing away data, the dataset is tailor made to be fit for purpose.

2.3. Analysis

During the analysis step, the data is used in order to answer the questions of the stakeholder(s) or the questions asked by the researchers themselves. A researcher has many algorithms, libraries, and methods to get these answers. In the case of Wi-Fi tracking, the usual analysis is on devices (or users, if that information is available), on characteristics of the sensor (the location), which both can give insight into activity (stay at locations) and identity (behavioural patterns) of a device or a user.

The analysis part of this project deals with the calculation of the number of users and the activities they might do during the irregular hours in the buildings on the TU Delft campus. The use of the spaces on the different APs is not given, therefore in order to assign an activity to a user at a specific location in the campus, a different method is used. A model is used in order to assign an occupation to a user (student, academic staff, support staff or other) and based on the location and some characteristic of the user, a probable activity is assigned. Box (1987),(2005) noted "All models are wrong but some are useful". The above aphorism was explained by Burnham and Anderson (2002) "A model is a simplification or approximation of reality and hence will not reflect all of reality...While a model can never be truth, a model might be ranked useful". Based on these ideas, this project is focused on building a model to assign users an occupation and then assign users with possible activities. Those models can be deterministic or stochastic. One model that can be used is the Hidden Markov Model (HMM), which is a stochastic model and can be applied to find patterns and underlying relations. It will be utilised to analyse the context of the dataset and discover patterns that represent different profiles of users. This can be done with the use of a training set of users with different profiles, such as a set of students, and a set of staff members. The training sets can come from any source, in this project they come from a questionnaire that is designed by the researchers. Depending on the behavioural patterns of a user and the comparison with the training sets, all the users of the dataset are assessed and the corresponding training set profile is assigned (Lara and Labrador, 2012). Thus, through the HMM the user's profiles of all the users of the campus will be determined. The model needs to be validated and applied, in order to verify the training sets of the user's profiles (Lin et al., 2015). The goal of the user profiling is to aggregate individual users to groups of users with common interests and requirements (TU Delft, 2013b). It is assumed that these groups also have similar behaviour. The activity of a user follows as the behaviour at a certain time and place. The assignment of user activities will be determined by a simple deterministic model based on assumptions.

2.4. Presentation of Results

The presentation of the data and of the results is an important step in the process of geomanaging. Tables with just numbers are meaningless, and can hardly be used to identify patterns that form the basis for understanding geo-phenomena and ultimately improving the (built) environment/world. Therefore, visualisation techniques should be applied to make the numbers meaningful.

There are different ways to visualise data and make it suitable for pattern identification. One way to visualise data is to highlight certain numbers in a table, for complex data this could be a first step. Another way to visualise data is to present it in graphs, either bar charts for discrete data, pie charts for statistical data, line charts for continuous data, and maps for geodata. Either options can be on paper, in digital sheets and maps, or interactive dashboards.

When presenting data the data should show patterns that relay a message, or tell a story. All of them could start on a large scale and narrow down and go into more detail into a smaller scale, or the other way around, start small and work towards a bigger picture. Different clients and users can identify patterns in their own way. The data specialist might see partners from raw numbers, others might easily read graphs, and some need a message or story to see the big picture. For all clients and users the appropriate technique of visualising data should be tailored to them specifically. The visualised results of this Wi-Fi tracking project are predominantly focused on answering the questions of the client FMRE, but will also provide valuable insight in how eduroam Wi-Fi logs are acquired and if the data hides issues related to privacy. The results are aiming to fulfil the scope of this project that is described in the MoSCoW Rules in section 1.6.

2.5. Quality Control

The importance of quality control is to make sure that during each step of the acquisition, processing, analysis, and presentation, the result is meaningful and valid; it should be sufficiently accurate, representative and not in violation of any legislation (i.e. an infraction upon privacy legislation). During the project, from start to finish, assumptions, aggregations, dis-aggregations and analyses are being carried out. In these processes, it is important that after each step no important information is thrown away. However, results also need verification and quality assessment in order to ensure that the final output is fit for use. The verification of the results is more oriented in comparing the outcome results with reality and deliver the related accuracy. Furthermore, assumptions should be backed up by common sense, reference data or literature. Quality control ensures that the result will be usable according to its stated accuracy.

3

Applied Methodology

In this chapter, the general methodology that is applied to the 'Rhythm of the Campus' project is broken down into parts and is discussed on a technical level. This chapter contains, firstly the data structure that is maintained in the database system, secondly the pre-processing of the data to make it suitable for analysis, thirdly the multiple analyses that are done on the data, fourthly the different types of presentation and visualisation are described, and lastly the quality of the workflow is assessed.

3.1. Data Structure

From the ICT department the Wi-Fi monitoring data is provided in the database system. On this system all the data is accessed, processed, and analysed using queries. The data structure as it exists within PostgreSQL is shown in Figure 3.1, with each table that is created during the queries. Furthermore, these tables reflect the milestone products and steps in the pre-processing and analysis.

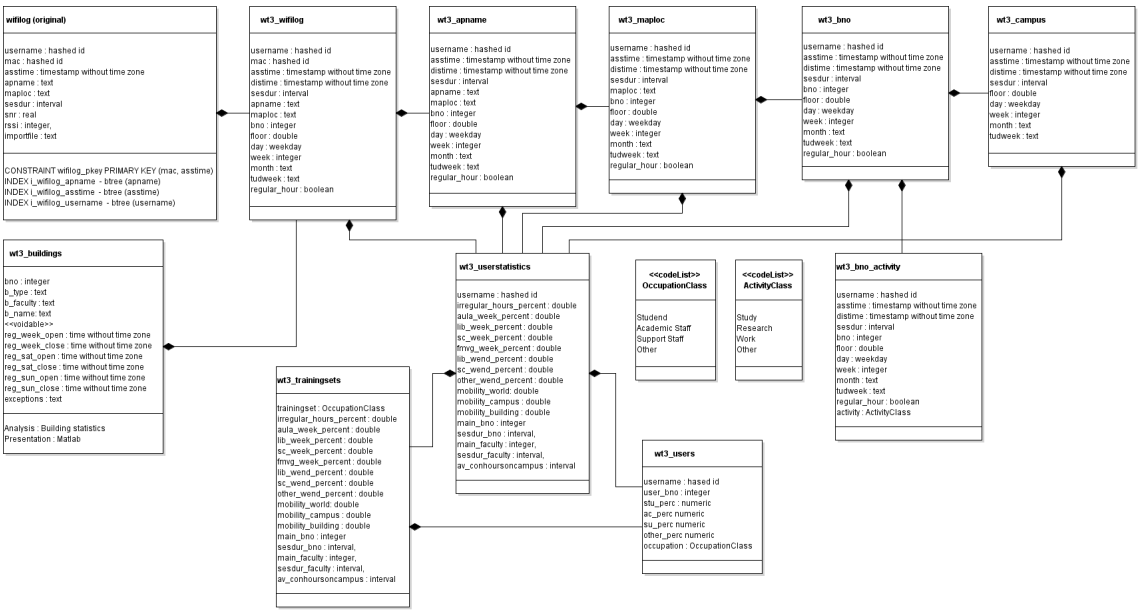


Figure 3.1: Class diagram with processing queries

3.1.1. wifilog table (acquired by ICT Department)

The starting dataset of this project is the dataset that is provided by the ICT department of TU Delft and contains all records straight from the source as a dump of each day. Each record consists of a hashed user (username), that is connected with his/her/its devices each with its unique MAC address (mac) with an access time (asstime) to a specific AP (apname) for a certain duration (sesdur) and with the average signal strength (rssi) and ignal to oise atio (snr). Each AP has a location on the campus/in a building that is reflected in the column maploc. Finally, each record originates from a dump of a specific day (importfile).

3.1.2. wifilog table (project wifilog)

The wt3_wifilog table is used for all the queries in analysis and visualisation. It has some columns removed, namely: rssi, snr and the column importfile. Since the spatial/temporal resolution is on building level, Wi-Fi fingerprinting, for the determination of the exact location in the building, is not within the scope of the research of this project.

During the data mining, the information that was implicitly contained in values, is extracted to explicit columns. From the apname column, the building number (bno) is extracted, and from the access time (asstime) column, the columns weekday (i.e. Monday), month (i.e. April), week number (i.e. 32) and TU Delft week as tudweek (i.e. 3.10) are extracted. Moreover, the column regular_hour is added, which signifies whether a session was started and finished while the corresponding building was closed. An example of the project wifilog is shown in Table 3.1.

Table 3.1: example of the project wifilog

user-name	mac	asstime	distime	sesdur	ap-name	map-loc	bno	day	week	month	tud-week	reg_h
User A	Mac A	2016-04-30 07:00:00	2016-04-30 09:00:00	02:00:00	A-23-0-035	….	23	satur-day	18	april	4.1	T
User A	Mac B	2016-04-30 09:31:00	2016-04-30 12:35:00	03:04:00	A-23-0-035	….	23	satur-day	18	april	4.1	T
User B	Mac B	2016-05-15 18:59:00	2016-05-15 20:50:00	01:51:00	A-40-0-035	….	40	sun-day	20	may	4.3	F

3.1.3. apname table

The wt3_apname table is an aggregate table of wt3_wifilog and contains all the information and columns of the wt3_wifilog, after the column ‘mac’ has been removed. This accounts to the fact that sessions of multiple devices of the same user are flattened (in the case of overlapping sessions) or stitched together (in the case of consecutive sessions). This table shows the time a user is connected (with any device) to the eduroam network, sessions on different APs are valid, meaning a user has left a device connected to one AP and moved in the building/campus to another AP with a second device. Mobility on the same floor could hypothetically be distinguished here.

3.1.4. maploc table

The wt3_maploc table is an aggregate table of the wt3_apname, where there is an aggregation from AP level to maploc level (each AP belongs to a certain floor or wing of a building). The apname column is no longer present after the aggregation. This table shows information about how long and how often a user was connected to a certain floor of a building. This allows the distinction of mobility between floors of a building.

3.1.5. bno

The wt3_bno table is an aggregate table of the wt3_maploc table, where the data is aggregated to building level, where the maploc is a subdivision of the building. The column 'maploc' has been removed. Currently, this table shows how long a user stayed within a building, without regarding the activities within the building. The only information contained is how often and for how long a user is in any location within a building. Overlap between buildings is allowed, meaning a user left one device on in a certain building and moved to another building (in itself valuable information).

3.1.6. campus table

The wt3_campus table is the highest level of aggregation from the wt3_wifilog, that shows how often and for how long a user is connected on the campus per day. With this table it is possible to identify when a user leaves campus (assuming that there is no connection to eduroam).

3.1.7. building table

The building table is composed of two parts: Initially, a static part that describes per building the characteristics of the building including name, abbreviation, opening hours during the week and weekend (not shown), and any exception to the opening hours. The information per building is found through the TU Delft website (TU Delft, 2016a). Additionally, a dynamic part, where the table is populated with the distinct number of people and the related percentage of people per building. Through the MoSCoW diagram in section 1.6, the described information answers to the first question that is related to the number of people that are present in each building during the irregular hours.

Table 3.2: Building table

bno	b_type	b_faculty	b_name	reg_week_open	exceptions
3	EDU		Science Centre Delft	08:00:00	
5	FAC	AS	Department of Biotechnology	07:30:00	authorized campuscard after closing
8	FAC	Arch	Faculty of Architecture and the Built Environment	07:00:00	friday 7:00 - 19:00
12	FAC	AS	Department of Chemical Engineering	07:30:00	authorized campuscard after closing
15	FAC	AS	TNW - Physical and Chemical Technology	08:00:00	
20	EDU		Aula Conference Centre	08:00:00	open during events (in weekend)
21	EDU		TU Delft Library	08:00:00	exam period: 8:00 am - 2:00 am
22	FAC	AS	Faculty of Applied Sciences	07:00:00	authorized campuscard after closing
23	FAC	CEM	Faculty of Civil Engineering and Geosciences	07:00:00	
30	FMRE		Education and Student Affairs/FMRE	07:30:00	
31	FAC	TPM	Faculty of Technology, Policy and Management	07:30:00	weekends 10:00 - 18:00 with authorized campuscard
32	FAC	IDE	Faculty of Industrial Design Engineering	08:00:00	7:00 - 8:00 with authorized campuscard, friday 8:00 - 19:00

Continuation of Table 3.2					
34	FAC	3ME	Faculty Mechanical, Maritime and Materials Engineering	08:00:00	authorized campuscard after closing
35	EDU		Education Building 35	07:30:00	
36	FAC	EEMCS	Faculty of Electronic Engineering, Mathematics and Computer Sciences	06:30:00	weekends 6:30 - 23:00 with authorized campuscard
37	SNC		Unit Sport	08:00:00	Sportcafe is open until 1:00 am
38	SNC		Unit Culture	08:00:00	
43	RES		Combined Heat and Power Plant		
45	RES		Low Speed Wind Tunnel Laboratory	08:00:00	research
46	RES		TNO	08:00:00	research
60	FMRE		Logistics and Environment	07:30:30	
61	RES		Delft Aerospace Structures and Materials Laboratory	07:00:00	
62	FAC	AE	Faculty of Aerospace Engineering	07:00:00	
63	RES		SIMONA Research Flight Simulator	08:00:00	research
64	RES		Aerodynamics Laboratory, Wind Tunnels	07:00:00	authorized campuscard after closing
66	EDU		The Fellowship	08:00:00	

3.1.8. user statistics table

The user table is a living table, where each query adds more information about the users. The main goal of the user table is to allow the classification of users into different profiles. With the columns of the table as input data, the HMM will classify every user as student, academic staff, support staff or other. The statistics are calculated from the sessions of the wt3_campus, wt3_bno and wt3_maploc tables.

3.1.9. training set table

This table contains only four registries which represent the typical behaviour of the four categories of user profiles. The four different user profiles are Student, Academic staff, Support staff, Other (e.g. guests). The typical behaviour of those user profiles come from the questionnaire (Appendix B and it serves to calculate the similarity of each user with those profiles. Hence, the profile table has the same data structure as the wt3_user statistics.

3.1.10. user table

The user table (as shown in Table 3.4) stores the main faculty or building of a user, the scores generated by the HMM of each user, and the actual profile assigned by HMM (student, academic staff, support staff, or other). This information can then be used to give a more in depth insight into the usage of (floor/wings of) buildings, beside the total count of users, the count can now be broken down into the four profiles.

Table 3.3: Example of the training set table.

username	condays_percent	conhours_average_percent	aula_rest_week_percent	lib_week_percent	sc_week_percent	FMRE_week_percent	lib_wend_percent	sc_wend_percent	other_wend_percent	mobility_world	mobility_campus	mobility_building
Student	0.4	0.16	0.05	0.15	0.1	0.001	0.2	0.2	0.08	0.55	0.60	0.80
Academic staff	0.63	0.35	0.005	0.001	0.08	0.0001	0.002	0.01	0.20	0.15	0.25	0.25
Support staff	0.63	0.33	0.04	0.003	0.05	0.5	0.01	0.05	0.01	0.1	0.15	0.2
Other	0.045	0.1	0.1	0.01	0.0005	0.0001	0.2	0.05	0.1	0.3	0.3	0.3

Table 3.4: User table sample

Username	user_buidling	occupation
User A	8	Student
User B	3	Support Staff

3.2. Pre-prossessing

Pre-processing is a procedure that takes place before the analysis of data and it aims to render the data suitable for further processing and analysis. In this part, cleaning of data, data mining and filtering data takes place. In Figure 3.2 the processes in pre-processing are shown arranged by the order of processing and dependencies.

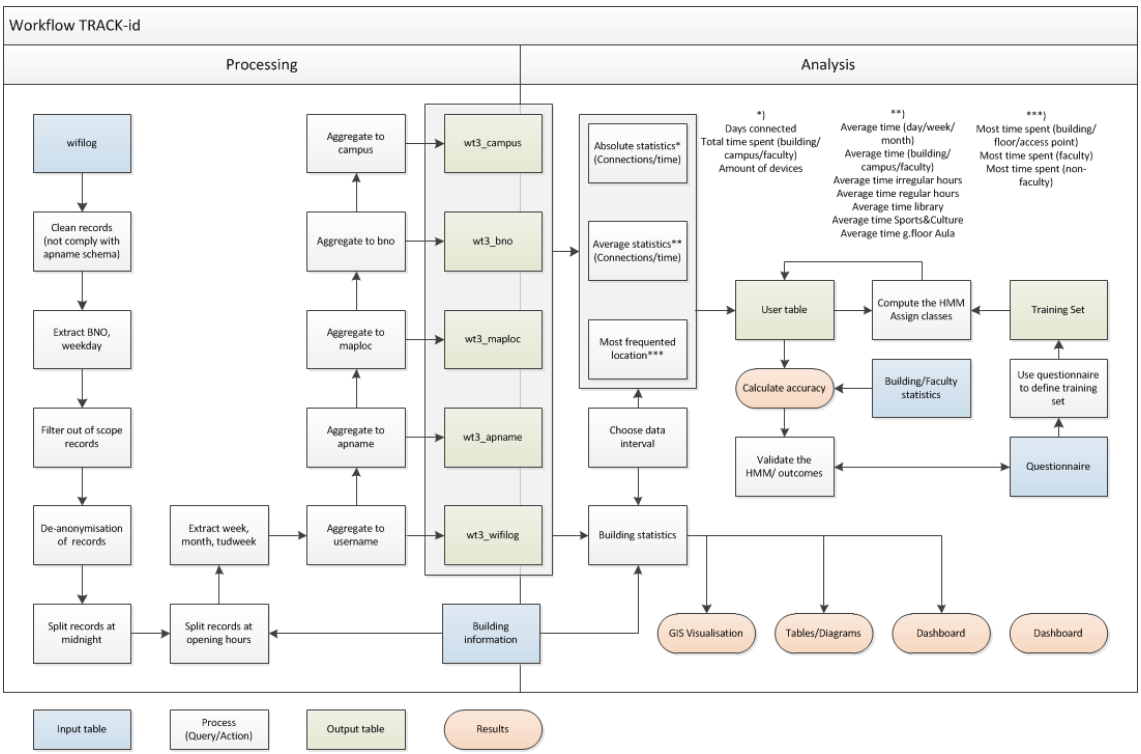


Figure 3.2: Workflow of queries arranged by order and dependency

3.2.1. Cleaning of data

Cleaning out of scope data

Since Wi-Fi fingerprinting is not within the scope of this research, the columns rssi, snr and importfile are dropped from the original wifilog table (shown in Table 3.5) and are not used in any subsequent step.

Table 3.5: Example of the original wifilog provided by ICT

	user-name	mac	asstime	apname	maploc	sesdur	snr	rssi	importfile
1	User A	Mac A.A	24-4-2016 23:45	A-08-B-102	08-BK-City > 1e Verdieping	00:25:08	30	-64	20160425-wifitracking.csv.gz
2	User A	Mac A.A	25-4-2016 00:15	A-08-B-102	08-BK-City > 1e Verdieping	01:21:30	38	-58	20160425-wifitracking.csv.gz
3	User A	Mac A.B	25-4-2016 07:11	A-08-E-004	08-BK-City > Beganegrond	00:30:00	18	-77	20160425-wifitracking.csv.gz
4	User B	Mac B.A	18-4-2016 15:11	A-08-H-003	08-BK-City > Beganegrond	01:07:33	14	-81	20160418-wifitracking.csv.gz
5	User C	Mac C.A	19-4-2016 17:55	A-12-0-012	12-TNW-DCT > Beganegrond	03:06:08	25	-69	20160419-wifitracking.csv.gz
6	User C	Mac C.B	19-4-2016 21:17	A-12-0-012	12-TNW-DCT > Beganegrond	00:20:31	46	-53	20160419-wifitracking.csv.gz

Data that does not comply with the schema

In order to start data mining and not make custom queries that have to take into account every deviation of the schema, records that don't comply with the schema, are removed. The AP name (apname) is consistent in most buildings, but a small amount of APs have a value that does not follow the usual schema. The records connected to those APs are cleaned, to make all consecutive queries run smoothly, with consistent apname naming. All AP names not starting with a 'A-' are being removed, see Table 3.6.

Table 3.6: Example of inconsistent and consistent AP naming

Inconsistent apname	Consistent apname
APe4aa.5db8.c0f0	A-21-0-044
N/A	A-08-E-104
APe4aa.5db8.bea4	A-26-0-012
AP5c83.8f89.71b4	A-22-0-030
TUvisitor	A-58-0-003
APe4aa.5dac.5668	A-30-0-015
APe4aa.5da0.1708	A-45-0-007
AP5c83.8f9f.d818	A-22-0-006
AP5c83.8f99.3678	A-134-0-039
AP54a2.74af.9660	A-21-0-032

Invalid session duration

Some records have an invalid session duration that results in consecutive sessions with an overlap period, where a device is connected to more than one AP. Based on the system of APs this is impossible (see section 4.3). Table 3.7 shows an example for one of this records. The first connection with a session duration of 09:18:59 hours is still ongoing, while a second connection with the same mac address is made around five minutes later. These records are identified as gross errors in the data.

Since the correct duration of the invalid session is not known the session is edited to end at the start time of the next session.

Table 3.7: Example of an invalid record, the same device has a duplicate connection

username	mac	asstime	apname	maploc	sesdur
User A	Mac A	2016-04-26 14:18:10	A-134-0-006	TUD-Collegezalen > 3ME-Collegezaal-C	09:18:59
User A	Mac A	2016-04-26 14:29:05	A-134-0-005	TUD-Collegezalen > 3ME-Collegezaal-C	00:05:16

3.2.2. Data mining

Extracting implicit information for pre-processing

In a first step of the data mining implicit information from the original wifilog (shown in Table 3.5) is extracted and a temporary wt3_wifilog (subsection 3.1.2) is created. The information that is extracted during this step is the Building Number (BNO) that is encapsulated in the apname, and the weekday. These two columns are needed when we split the records at opening hours.

Filter out of scope records

The scope of the research does not include all buildings. Only records that belong to the buildings within the scope are kept, records that are out of the scope are filtered out and not kept for further analysis.

De-anonymise anonymous@tudelft.nl

The user with the most devices is anonymous@tudelft.nl, it's the anonymous username that is used in the TTLS authentication protocol of eduroam. A large part of the device sessions of user anonymous@tudelft.nl can be de-anonymised and can have the not-anonymous username assigned to the device instead of anonymous, the username that is assigned can be obtained by finding the other username that also uses the device frequently (non-anonymous user). The sessions that remain (have no alternative non-anonymous user) are filtered out.

Splitting of records to improve statistics

Records that span over more than one day are split up at midnight to separate the connections per day. The session duration is cut at midnight and the disconnection time is set to '24:00:00'. The rest of the session duration is added in a new record on the next day, or recursively split again if it passes another day. Splitting the records at midnight ensures that these columns always represent the true values of the session when doing statistics about days/weeks/months.

Additionally, to the splitting at midnight hours, every record is checked for overlap with an opening or closing hour that is taken from the building table according to the extracted BNO of each session. If a record spans over an opening/closing hour, it is split up in a regular part (inside opening hour) and an irregular part (outside opening hour) in the same way as the splitting is done at midnight. From this, a Boolean value is assigned to the regular_hour column that is true for the dataset which is related with the regular hours and false for the dataset, which is related with the irregular hours.

The resulting table is shown in Table 3.8. Row 2 and 3 show an example of a record that was split at midnight. The record in row 5 in the original wifilog table spans over '18:00' o'clock that is the closing hour of the Chemical Engineering building. Therefore, the record is split and results in the rows 7 and 8, in Table 3.8.

Table 3.8: Wifilog_t3 with explicit information about building number and dates

user-name	mac	asstime	distime	sesdur	ap-name	map-loc	bno	day	month	week	tud-week	reg_h
User A	Mac A.A	24-4-2016 23:45:07	2016-4-24 24:00:00	00:14:53	A-08-B-102	...	8	Sun-day	April	16	4.1	f
User A	Mac A.A	25-4-2016 00:00:00	25-4-2016 00:10:15	00:10:15	A-08-B-102	...	8	Mon-day	April	17	4.2	f
User A	Mac A.A	25-4-2016 00:15:15	25-4-2016 01:36:45	01:21:30	A-08-B-103	...	8	Mon-day	April	17	4.2	f
User A	Mac A.B	25-4-2016 07:11:46	25-4-2016 07:41:46	00:30:00	A-08-E-004	...	8	Mon-day	April	17	4.2	f
User B	Mac B.A	18-4-2016 15:11:38	18-4-2016 16:55:11	01:43:33	A-08-H-001	...	8	Mon-day	April	16	4.1	t
User C	Mac C.A	19-4-2016 17:55:07	19-4-2016 18:00:00	00:04:53	A-12-0-012	...	8	Tues-day	April	16	4.1	t
User C	Mac C.A	19-4-2016 18:00:00	19-4-2016 21:01:15	03:01:15	A-12-0-012	...	8	Tues-day	April	16	4.1	f
User C	Mac C.B	19-4-2016 21:17:55	19-4-2016 21:38:26	00:20:31	A-12-0-012	...	8	Tues-day	April	16	4.1	f

Extracting implicit information for analysis

After the splitting the wt3_wifilog is finalised by adding the columns 'week' and 'month' with built-in PostgreSQL functions where the asstime is used as input. The TUD week is manually added with a case function.

3.2.3. Aggregation

Pre-processing of the data is needed, in order to meet the client's requirements and answer their questions. The writing of the queries involves the aggregation of the data of the dataset to account for overlap and continuous sessions, this result is stored in subsequently more aggregated tables (wt3_apname, wt3_maploc, wt3_bno and wt3_campus). As described, the wt3_wifilog provides all the connections made by a user and all of his/her/its (different) devices to the eduroam network of the TU Delft. For profiling the user in a later stage of the analysis, the distinction between devices is of no interest, because the scope of our research is on the profiling of the users and not on the profiling of the different devices of a user. Therefore, the different sessions of each MAC address/device of a user is aggregated. The resulting tables have the continuing sessions of users per AP, Maploc, Building, or Campus.

Aggregation to building

To illustrate the process of aggregation, Table 3.9 shows part of the data in the wt3_wifilog for a certain user on a certain day (18-04-2016). For this user there are 10 records which show that the user in the morning was at the faculty of Architecture, in the afternoon again at the faculty of Architecture and in the evening he/she visited the Library. This makes it clear that there will be many sessions for a certain user in a certain building. A script is created to aggregate all these sessions for each user from the moment the user gets connected to eduroam, until he/she get 'disconnected' in a certain building. The resulting wt3_bno for this specific user is shown in Table 3.10.

3.2.4. Subset of the total data

In order to answer the requested questions (section 1.6), only records that are relevant for the answer should be considered. A subset of the wt3_wifilog, wt3_apname, wt3_maploc, wt3_bno, or wt3_campus can be created and used to answer questions of the client.

Table 3.9: Wifilog_t3 for User A on the 18th of April 2016

	username	mac	asstime	distime	sesdur	apname	maploc	bno
1	User A	Mac A.A	18-4-2016 09:45:07	18-4-2016 09:50:15	00:05:08	A-08-C-202	...	8
2	User A	Mac A.A	18-4-2016 09:50:15	18-4-2016 13:11:45	03:21:30	A-08-B-102	...	8
3	User A	Mac A.A	18-4-2016 13:11:46	18-4-2016 13:16:46	00:05:00	A-08-E-004	...	8
4	User A	Mac A.B	18-4-2016 15:11:38	18-4-2016 16:19:11	01:07:33	A-08-H-001	...	8
5	User A	Mac A.A	18-4-2016 15:11:38	18-4-2016 16:55:11	01:43:33	A-08-H-001	...	8
6	User A	Mac A.B	18-4-2016 16:29:32	18-4-2016 16:50:03	00:20:31	A-08-H-001	...	8
7	User A	Mac A.B	18-4-2016 20:57:43	18-4-2016 21:32:52	00:35:09	A-21-0-006	...	21
8	User A	Mac A.A	18-4-2016 20:57:43	18-4-2016 21:02:44	00:05:01	A-21-0-063	...	21
9	User A	Mac A.A	18-4-2016 21:02:44	18-4-2016 21:57:58	00:55:14	A-21-0-042	...	21
10	User A	Mac A.B	18-4-2016 21:32:52	18-4-2016 21:57:56	00:25:04	A-21-0-063	...	21

Table 3.10: Activitylog for User A on the 18th of April 2016

	username		asstime	distime	sesdur			bno
1	User A		18-4-2016 09:45:07	18-4-2016 13:16:46	03:31:39			8
2	User A		18-4-2016 15:11:38	18-4-2016 16:55:11	01:43:33			8
3	User A		18-4-2016 20:57:43	18-4-2016 21:57:58	01:00:15			21

Selecting a subset

The dataset that is used, is big and may include dataset of many days, weeks, months etc. Therefore, an automatic technique is created to provide the clients a subset of the dataset depending on the period of time that they want to analyse. More specifically, the client can choose a dataset depending on a building number (integer), the weekend (Boolean), the weekdays (Boolean), a day (text), a week (text), a month (text) and the regular or irregular hours (Boolean) or a combination of them. With this technique, errors of the choice of a subset of the dataset are eliminated and a comparison of different subsets of the dataset is provided in less time.

3.3. Analysis

During the analysis step the more complex data enrichment techniques and statistics are used to work towards the answers of the client. The analysis that is linked with the buildings of the campus of TU Delft is related with the number of people that get access to the Wi-Fi through the different buildings of the campus. By analysing the different data from different, non-overlapping time intervals interesting information can be extracted.

3.3.1. User statistics

First the user statistics table is created. This table contains statistical information on every individual user of the wifilog. The table is used to determine the occupation of the users with the help of the HMM. The column condays_percent of the table shows the percentage of days a user is connected. The column conhours_average_percent shows the average number of hours a user spends on campus, relative to one day. The following seven columns show the number of hours a user spends in the Aula restaurant, Library, Sports and Culture or FMRE building, relative to the total number of hours he spends on campus. Furthermore, weekdays, Monday to Friday, are distinguished from weekends. In the second part of the table the mobility columns show the average connections related to the total connection days as an inverse probability.

Assigning a building/faculty to a user

The main_bno and main_faculty column show the building number of the faculty a user spends the most time at. The preliminary way to tag a user's faculty depends on the building in which he or she spends the greatest proportion of time. Since this can also be a non-faculty building as the library or aula, not every user can be assigned to a home faculty in this way. The Main_faculty column therefore shows the faculty (or FMRE building) a user spends the most time at. There are a total of 11700 users that spent more time at an Education or Research building than in a Faculty. Most of those users should later be classified as Academic, Support Staff or Other (e.g. guests). For the activity statistics (subsection 3.3.3) only the main building of people is used.

Table 3.11: Example of user statistics

username	condays_percent	conhours_average_percent	aula_rest_week_percent	lib_week_percent	sc_week_percent	FMRE_week_percent	lib_wend_percent	sc_wend_percent	other_wend_percent	mobility_world	mobility_campus	mobility_building	main_building	main_faculty
User A	0.4	0.16	0.05	0.15	0.1	0.001	0.2	0.2	0.08	0.55	0.60	0.80	21	8
User B	0.63	0.35	0.005	0.001	0.08	0.0001	0.002	0.01	0.20	0.15	0.25	0.25	21	8
User C	0.63	0.33	0.04	0.003	0.05	0.5	0.01	0.05	0.01	0.1	0.15	0.2	21	8
User D	0.045	0.1	0.1	0.01	0.0005	0.0001	0.2	0.05	0.1	0.3	0.3	0.3	21	8

3.3.2. Profiling using the HMM

In order to distinguish the activity that a user does, different profiles are assigned to the users. In this part, the profile of the users are distinguished based on the HMM. Some training sets are defined considering the different profiles that need to be assigned to the users, such as student, academic staff, support staff or other. Those training sets are compared with the user's information that are derived from the dataset.

A Markov model is a stochastic model used to model randomly changing systems, where it is assumed that future states depend only on the current state and not on the events that occurred before it. The stated property is characterised as 'memorylessness' or Markov property. Generally, this assumption enables reasoning and computation with the model that would otherwise be intractable. There are different Markov models used in different situations. The simplest Markov model is the Markov chain. It models the state of a system with a random variable that changes through time. A Hidden Markov Model (HMM) is a Markov chain for which the state is only partially observable. In other words, observations are related to the state of the system, but they are typically insufficient to precisely determine the state. A Markov decision process, is a Markov chain, in which state transitions depend on the current state and an action vector that is applied to the system. A Markov random field, or Markov network, may be considered to be a generalization of a Markov chain in multiple dimensions. In a Markov chain, state depends only on the previous state in time, whereas in a Markov random field, each state depends on its neighbors in any of multiple directions. Hierarchical Markov Models can be applied to categorize human behavior at various levels of abstraction (Figure 3.3). For example, a series of simple observations, such as a person's location in a room, can be interpreted to determine more complex information, such as in what task or activity the person is performing (Petrushin, 2000), (Lühr et al., 2003), (Mühlenbrock et al., 2004), (Stamp, 2015).

In this project case, there is only one state transition, from an unknown occupancy state to the occupancy state, the model is utilised to determine the occupation of each individual that is identified in the

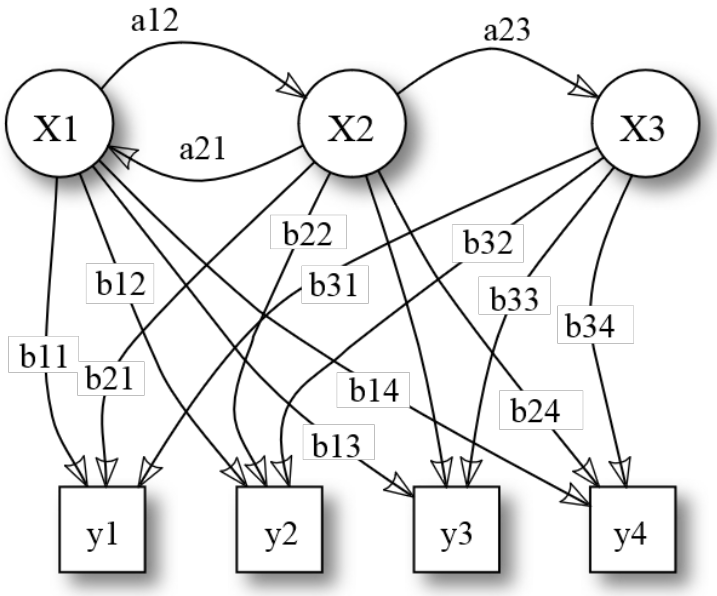


Figure 3.3: Visualisation of Markov Model (Wikipedia, 2012)

TU Delft buildings during the irregular hours. The occupation along with the location (building level) are the determining factors to assume the user's activity in the TU Delft buildings. The model is used in the identification of occupation using multiple factors and tried to connect the presence of oneself to certain places and time with its occupation. The connection is made using training sets for each occupation category and a probability for each user and each occupation category has been calculated. Each individual is assigned to an occupation category randomly, using its occupation probabilities.

A questionnaire has been designed to collect the information with which the training sets of each occupation category are created. The anonymous questionnaire aims to collect information that can also be derived from the dataset of this project for each user. The questionnaire also serves as an informing document of the procedure that takes place and as a mean to gather weak consent of the subjects. The questionnaire can be found in the Appendix B.

The project questionnaires could not be distributed, hence the training sets were based on assumptions. The training sets serve to compare the similarity between each user and the different occupations. Based on the similarity, the occupation probabilities are calculated for each user. The training sets are composed of eleven factors that describe the average behaviour of the different occupation categories. Those factors are supposed to differ significantly for each occupation category. Moreover, they should describe the typical behaviour of each occupation.

After the similarity between each user and the training sets has been calculated the probability for each occupation that a user has can be derived and based on them an occupation can be assigned to each user. However, the Markov model is a stochastic model which means that even if an occupation is the most probable for a user, the model is going to assign him an occupation randomly, based on the probabilities.

3.3.3. Activity assumptions for different occupations and locations

Finally, from user occupation, user activities are determined using some assumptions. Those assumptions take into consideration the user's profile, including his main building (subsection 3.3.2), and the

location where the user was tracked to determine his activity. For example, when a user is assigned the academic staff occupation and is located in his faculty building, then the user activity is set to research. In this part we assign a specific activity to every session of the wt3_bno table. All the assumptions for the four different activities Study, Research, Work and Other are given in Table 3.12 below. The outcomes are given in appendix A.

Table 3.12: Activity assumptions for different occupations and locations

	Location		Activity			
	Type	Main building	Study	Research	Work	Other
Student	RES	irrelevant		x		
	FAC	irrelevant	x			
	EDU	irrelevant	x			
	SNC	irrelevant				x
	FMRE	irrelevant				x
Academic Staff	RES	irrelevant		x		
	FAC	yes		x		
		no			x	
	EDU	irrelevant	x			
	SNC	irrelevant				x
	FMRE	yes			x	
		no				x
Support Staff	RES	yes			x	
		no				x
	FAC	yes			x	
		no				x
	EDU	yes			x	
		no				x
	SNC	yes			x	
		no				x
	FMRE	yes			x	
		no			x	
Other Staff	RES	irrelevant		x		
	FAC	irrelevant	x			
	EDU	irrelevant	x			
	SNC	irrelevant				x
	FMRE	irrelevant				x

3.3.4. Building statistics

The extent of use of different buildings is analysed in two ways. First with the amount of unique users connected to a certain building. This data can easily be collected from the wt3_bno table. Additionally the amount of time for all the users spent during a certain period is calculated for each building. Both datasets are examined for the regular and the irregular hours, during different periods of time through the given dataset. Thereby, the difference in use during opening and closing hours is shown. The count of users is split into the different profiles and the time spent in a building is split into different activities. This will give a clear insight into the use of the campus per building, an example of the analysis in amount of users is is shown in Table 3.13, an example of the analysis on time is shown in Table 3.14.

Table 3.13: Example of the amount of users on campus, subdivided by profile

Regular Hour	Profile	4-1-2016	4-2-2016	4-3-2016	4-4-2016
TRUE	TOTAL	15531	2606	2378	15824
FALSE	TOTAL	1725	2715	2975	2477
TRUE	Academic Staff	2388	357	314	2420
TRUE	Other	1267	247	214	1294
TRUE	Student	10005	1721	1606	10250
TRUE	Support Staff	1871	281	244	1860
FALSE	Academic Staff	275	460	488	406
FALSE	Other	113	215	249	189
FALSE	Student	1127	1719	1893	1588
FALSE	Support Staff	210	321	345	294

Table 3.14: Example of the time spent on campus subdivided into activities

Regular Hour	Activity	4/1/2016	4/2/2016	4/3/2016	4/4/2016
TRUE	TOTAL	74401	13202	11306	83878
FALSE	TOTAL	2271.1	8840.1	10580	3593
TRUE	Other	3642.9	968.88	913.63	4502.2
TRUE	Research	8578	247.5	24.822	9371.1
TRUE	Study	53736	10952	9806.8	60823
TRUE	Work	8444	1033.7	560.38	9181.6
FALSE	Other	139.62	536.13	693.75	188.59
FALSE	Research	337.6	836.9	718.68	486.31
FALSE	Study	1504.8	6658.6	8327.5	2492.6
FALSE	Work	289.13	808.49	840.38	425.48

Additionally, the count of user is also analysed in more spatial and temporal detail. By counting users per maploc for every hour in the dataset, certain floors or areas of interest can be determined.

Some unusual activities can be detected throughout the dataset. The procedure starts with the number of distinct users per building per day, which represents how significantly the day contributes to the whole irregular behaviour in this building. Examples are shown in findings (section 3.5).

3.3.5. Map location statistics

Each session record is classified into one time slot, for example, asstime = 00:05:00, distime = 00:35:00 then the session record will be classified into time slot 00:00:00 to 01:00:00. Aggregate the original wifilog data with map location, map location, week, day and time slot and then count distinct users in one time slot. Take the case shown in following table, only 2 users were at the ground floor of the Faculty of Architecture and Built Environment for longer than six minutes on the Sunday of week 19 during 22 pm to 23 pm, an example is shown in Table 3.15.

Table 3.15: Example of maploc statistics

Bno	Maploc	Week	Day	User count	Time slot
8	First floor	19	Sunday	2	22:00-23:00

The identification of unsafe areas during irregular hours is also required by FMRE due to security

issues; the definition of an unsafe area is presumed as the map location with less than three users in one irregular time slot (when the attribute “regular_hour” = “FALSE”) in a one day. Therefore, the unsafe records (each record represents one independent time slot) are counted per each map location and then averaged by total days of the data. The results represent that how many hours per map location per day are unsafe irregular hours.
The top five map location with longest unsafe irregular hours are listed in Table 3.16.

Table 3.16: Top five map location with longest unsafe irregular hours

Faculty	Map location	Unsafe irregular hours/day
SIMONA Research Flight Simulator	ground floor	7.42
Faculty of Architecture and the Built Environment	1st floor	7.33
Faculty of Industrial Design Engineering	3rd floor	6.73
Electronic Engineering, Mathematics and Computer Sciences	9,10,11 floor	6.17
Faculty of Industrial Design Engineering	4th floor	6.07

On average, there are 7.42 hours per day, in each hour, less than three users are on the ground floor of SIMONA Research Flight Simulator Laboratory which might be caused by students and academic staff working in the laboratory.
There are 7.33 hours per day, in each hour, less than three users are on the first floor of Faculty of Architecture and the Built Environment which might be caused by students and academic staff studying in the faculty library.

3.4. Presentation

In this chapter the different visualisation techniques that are used to make the data readable and interpretable are discussed, namely: tables and graphs, a dashboard web application, and the GIS component of the dashboard.

3.4.1. Tables and graphs

After analysis the tables itself need to be visualised in order to see patterns and let the numbers speak for themselves. In order to get insight into the use during irregular hours, the visualisation must comply with a couple of criteria: it is necessary to give a context to the irregular hours, hence compare them with the regular hours (criteria 1, context) and there is the need to identify patterns through time for different days and/or weeks (criteria 2, temporal dimension), then the amount of both users and time are to be visualised (criteria 3, count) and who the users were and what they did during time spent (criteria 4, subdivision). All these 4 criteria are necessary to answer the questions of FMRE. These questions can be broken down into a general who did what when and where? The resulting visualisation is a mirrored stacked bar chart for both the amount of users and the amount of time separately per building/campus, that visualises both the irregular hours and regular hours on the Y-axis and time on the X-axis.

An example of the aforementioned visualisation is shown in Figure 3.4. There is a lot of information contained in Figure 3.4: for a specific location (building or campus) it shows per day the amount of users that were active in regular and/or irregular hours, the amount shown as numbers in the bar chart is the number calculated from the wt3_bno, this is as accurate as the original wifilog is, it is just a count for each day. The subdivision that is visible within each bar is the amount of the four different users, classified as students, academic staff, support staff, or other (respectively from dark to light). This subdivision is based on the HMM and contains assumptions and should be used as a guiding ratio, rather than hard numbers. On the X-axis the dates of the wifilog are shown with a tick at the starting day and a tick on each Monday, to separate the weeks. This chart should be used together with the national holiday calendar and the academic calendar to discover the patterns that are masked.

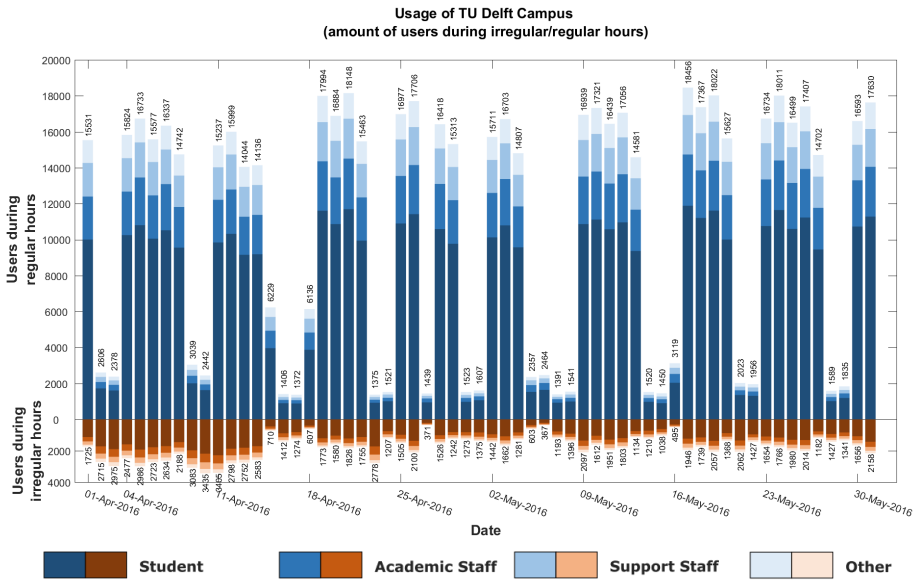


Figure 3.4: Visualisation of the amount of users on campus, subdivided into profiles

3.4.2. Web application

A dashboard is developed for easier visualisation and helping those who are not GIS experts. The dashboard consists with a map canvas and one functional sidebar as shown in Figure 3.5 which can on the fly generate charts by clicking and selecting. The dashboard user manual is included in the Appendix D.

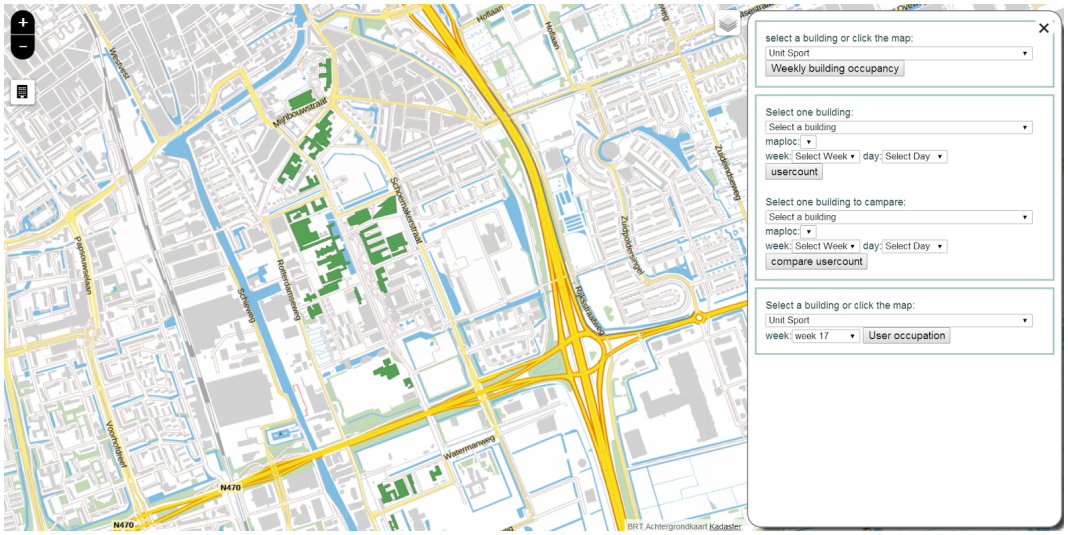


Figure 3.5: User interface

The dashboard has three main functions:
1. Stacked column shows the user count and occupation per week
The application can on the fly generate stacked column charts to show user count and user occupation (students, academic staff, support staff or other) per week (from week 13 to week 21) of a selected building during regular and irregular hours. Example of the Faculty of Architecture and Built Environment is shown in Figure 3.6.

2. Column chart represents the hourly occupancy of a specific floor of one selected building
The function counts user number per hour in a specified day at either map location or building level. The

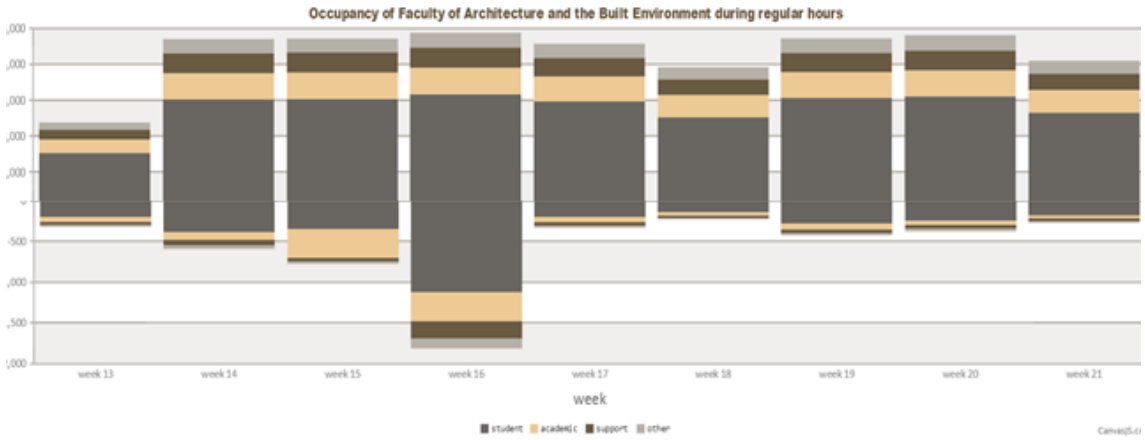


Figure 3.6: Building occupancy and user occupation during regular and irregular hours

change of occupancy of the building or map location along with the time series can be easily obtained and clearly visualised. Activities and events can be presumed according to the change of occupancy or the comparison of occupancy.
A sudden change of user number compared with the usual number also indicates an event. Example shown in Figure 3.7 represents the user number in Aula during a common Wednesday compared with what during King's day (Wednesday of week 17). A sharp decrease from 600 users per hour to 10 users per hour demonstrates that most users chose to spend King's day out of campus.

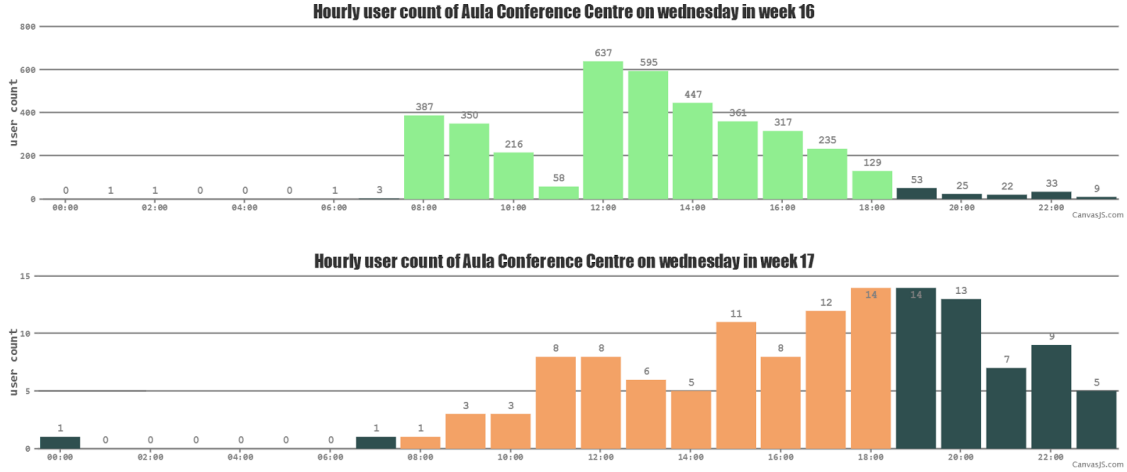


Figure 3.7: User count of Aula on Wednesday in week 16 and week 17

The example shown in Figure 3.8 represents the user number change during BK-Beats on the ground floor of Faculty of Architecture and Built Environment. It can be indicated that the event started at midnight of Friday in week 16 and ended at about 5 am the next day. The maximum number of the participator is around 800. About 500 people left between 4 am to 5 am because the end of music show. During the daytime of the Saturday in week 16, there were about 15 people still at the same map location; those people can be presumed as staff or students who are in responsible for cleaning the event space.

What is more, the habit of a user group can be presumed. Example shown in Figure 3.9 represents the user count per hour in the Unit Sports on Monday and Saturday of week 18. It can be found that, compared with the relative steady user count during Saturday; the peak of using the Unit Sports during Monday shifts backwards to between 18 pm and19 pm.

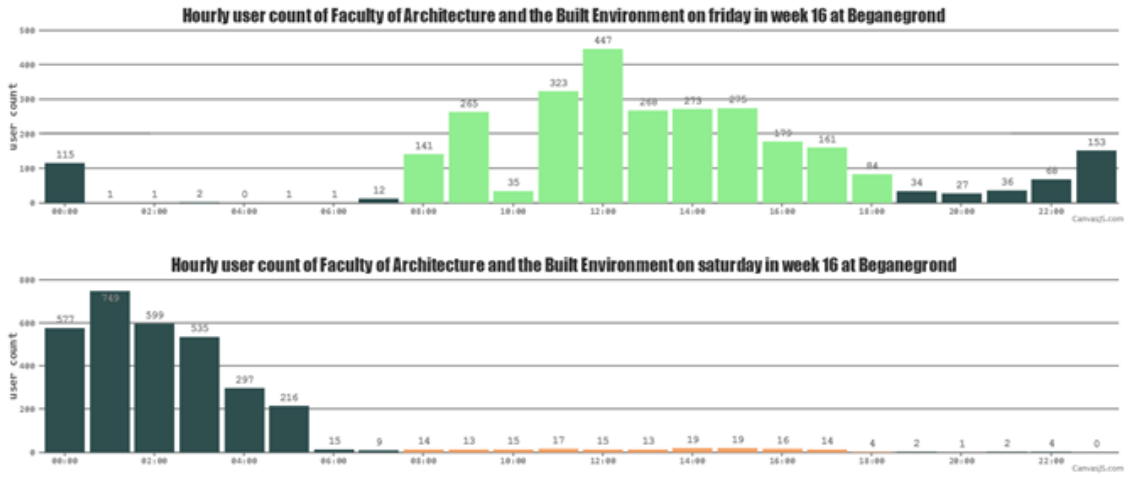


Figure 3.8: User count on the ground floor of BK in Friday and Saturday of week 16

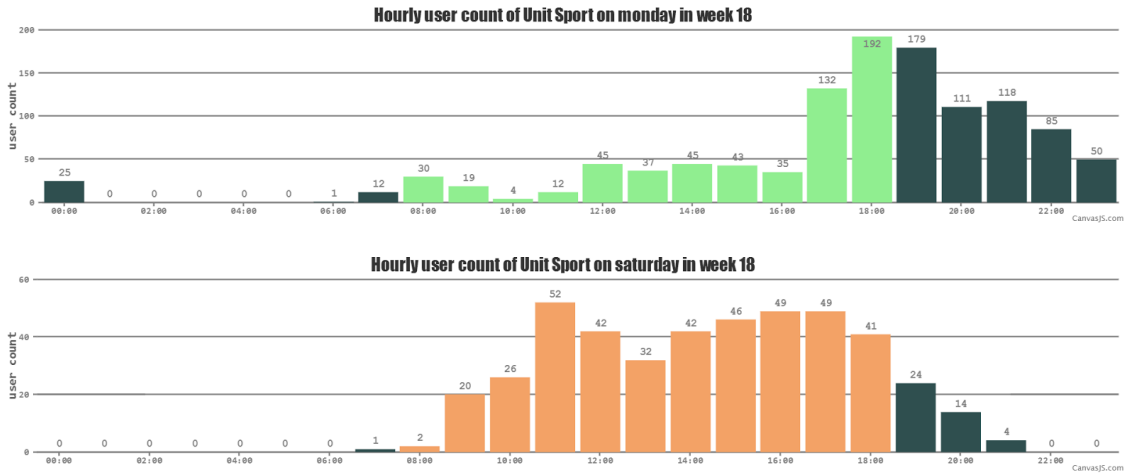


Figure 3.9: User count of the Unit Sports in Monday and Saturday of week 18

3. User group information who uses the selected building in a specified week
This function shows the user group of a building in addition with what are the users' main building and what are the usergroups' occupation. An example of information of the user group of the Unit Sports in week 15 is shown in Figure 3.10. It can be indicated from the Figure 3.10 that except the users whose main building is the Unit Sports, students from 3ME spend most time in the Unit Sports; academic staff from EEMCS spends the second most time there. What is more, the proportion of support staff in Unit Sports is larger than other buildings.

Example shown in Figure 3.11 represents the information of user group who uses the aerospace structures and materials laboratory in week 18. It can be found that except the people assigned to the building, people from the Faculty of Aerospace Engineering use the laboratory most in week 18. The result can be confirmed according to what is shown in Figure 3.12, in week 14, users from aerospace engineering spend 458 hours in the laboratory in total which is 300% of the time users from the laboratory spend here. What is more, the sum of time that users from the other buildings spend in the laboratory is only 55 hours. Therefore, the conclusion can be made that the main user group of the aerospace structures and materials laboratory is user from the Faculty of Aerospace Engineering.

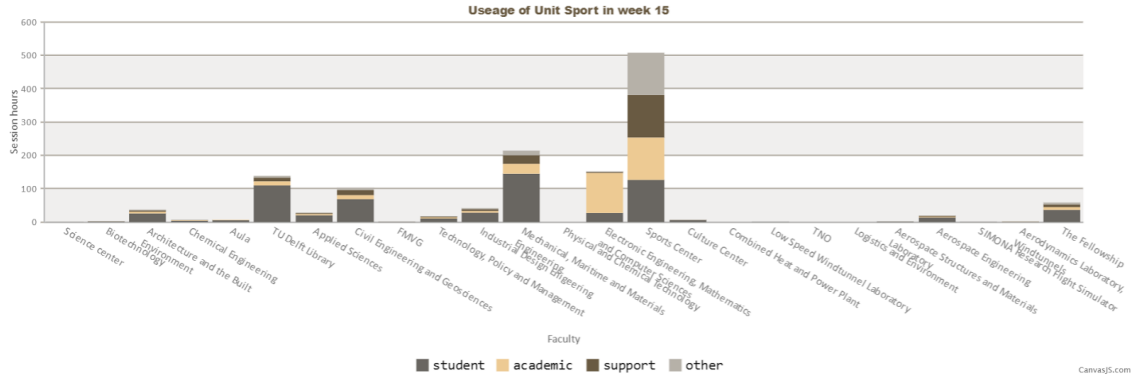


Figure 3.10: Information of user group who uses the Unit Sports in week 15

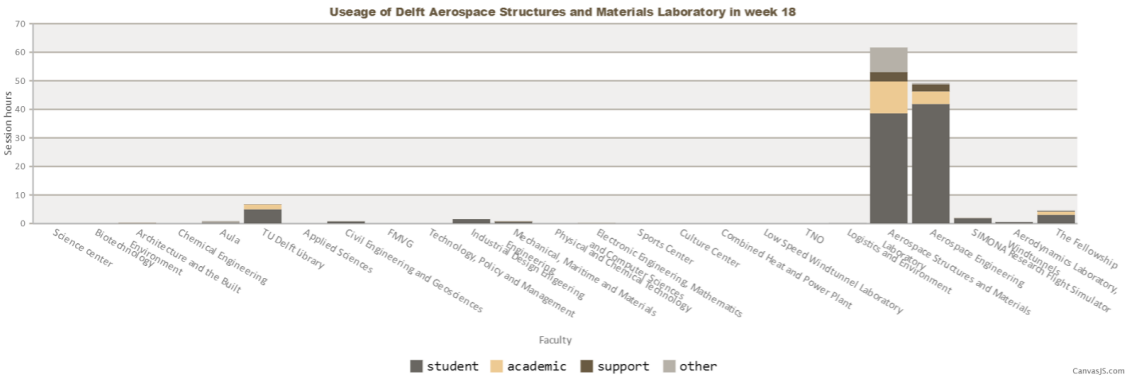


Figure 3.11: Information of user group who uses the aerospace structures and materials laboratory in week 18

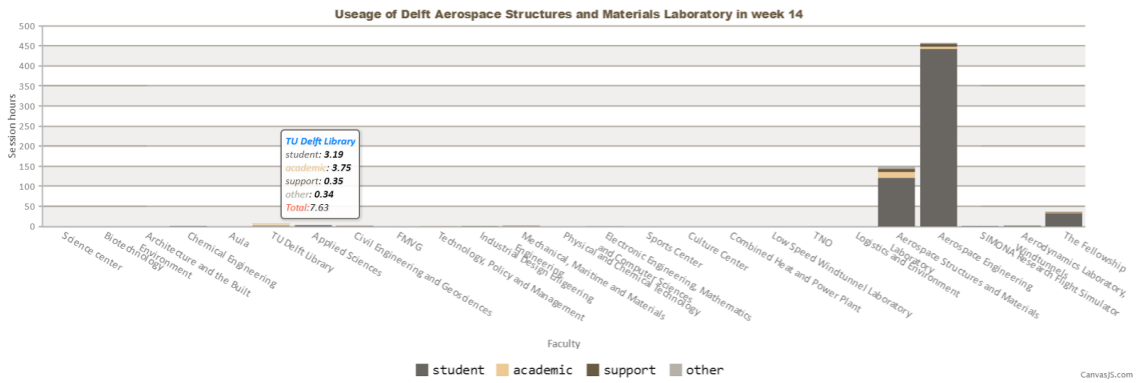


Figure 3.12: User Information about aerospace structures and materials laboratory in week 14

3.5. Findings

The following figures give insight into the amount of users on campus over the whole dataset. Figure 3.13 shows that more than 8000 users only have connections on one out of the 61 days in the dataset. The amount of users gradually decreases with the amount of connection days.

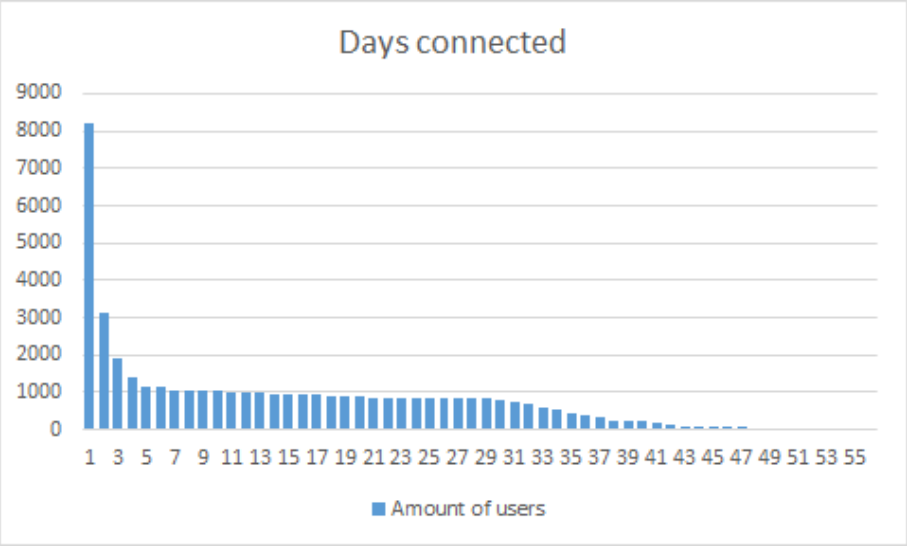


Figure 3.13: Count of (unique) users over the presence on campus

A similar pattern can be detected when plotting the amount of users over the total time spend on campus (see Figure 3.14). More than 5000 user spend up to one hour on the campus. An aggregated session duration up to two hours was only spent by around 2000 users. The rest of the users spent two or more hours on campus. There are also “users” that were connected for more than 1300 hours, which results in an average session duration of more than 23 hours per day. These “users” can be regarded as non-human users.

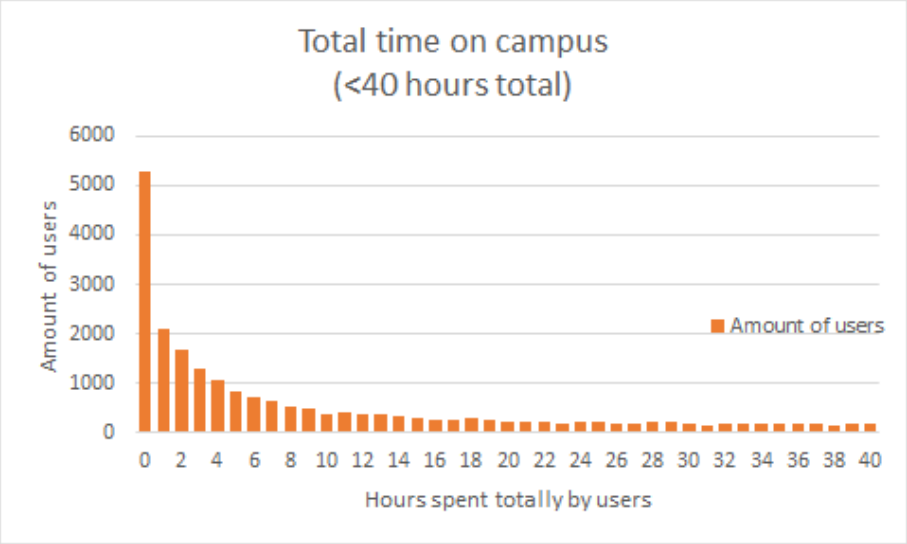


Figure 3.14: Count of (unique) users over the total time on campus

The average connection duration per day follows from the two Figures above (see Figure 3.15). As expected the biggest amount of users was connected only up to one hour. Besides that the most users have an average session duration between four and five hours.

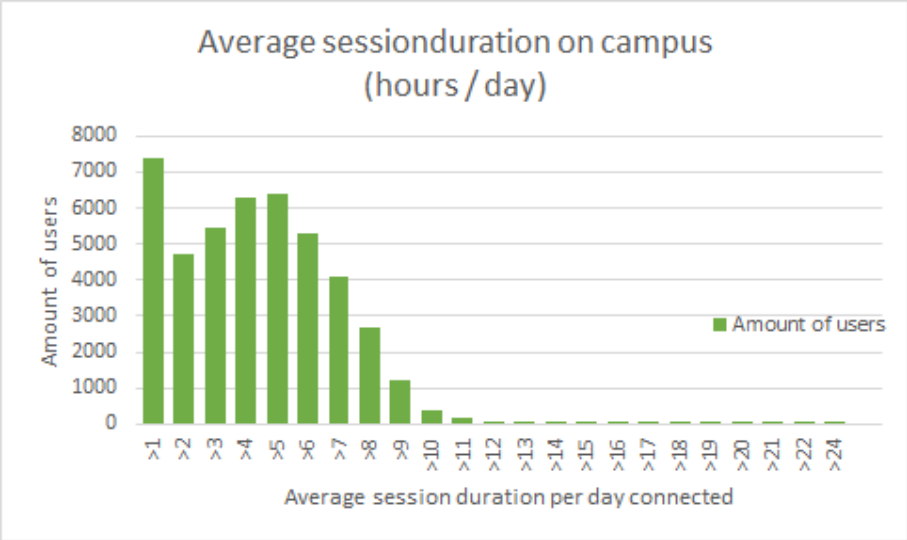


Figure 3.15: Count of (unique) users over the total time on campus

As illustrated in the Figure 3.16, during irregular hours there were around 70.000 students recognised, 17.500 Academic staff, 13.000 Support staff and 8.000 other users. It is regarded that the student number is high, because while counting users during irregular hours, the exceptions in buildings were not taken into consideration. Given the fact that the first two weeks were exam period in TU Delft and some buildings stay open for extended periods, there were many students that were counted to be there as in irregular hours while actually the buildings were still open. Additionally, as it is depicted in Figure 3.17, the approximate time each activity took place during irregular hours is around 131.000 hours of study, almost 28.000 hours of research, 23.000 hours of work and 11.000hours of other activities. Again, the study hours are overestimated because of the false exceptions. If the days until the end of exams are not taken into consideration, which is illustrated with a grey line in both Figures, 24.6% of days (fifteen out of sixty-one days) are removed and each occupation category during irregular hours drops around 36%, which is the first indication implying that those days had more users than usual during irregular hours. Moreover, by dropping those days, research and work activities drop by 31% and 36% accordingly while study and other activities drop by 47% and 48% accordingly. In other words, almost the 25% of days hold close to 50% hours spent on those activities during irregular hours.

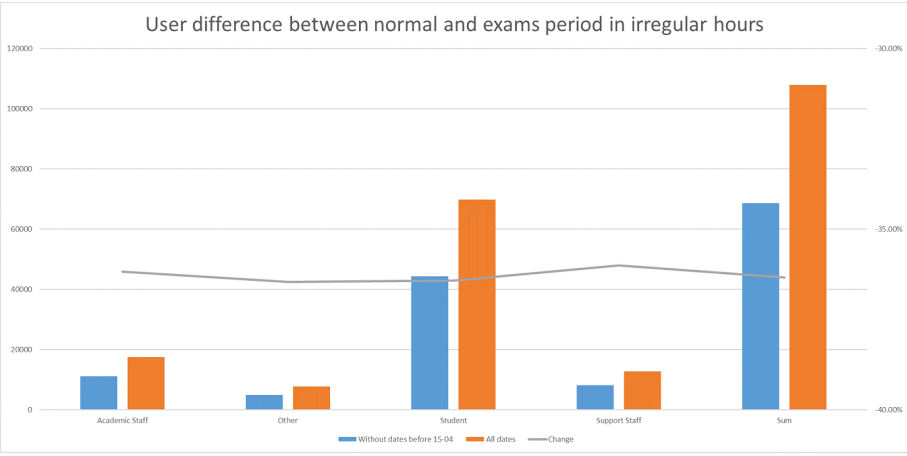


Figure 3.16: User difference between normal and exams period in irregular hours

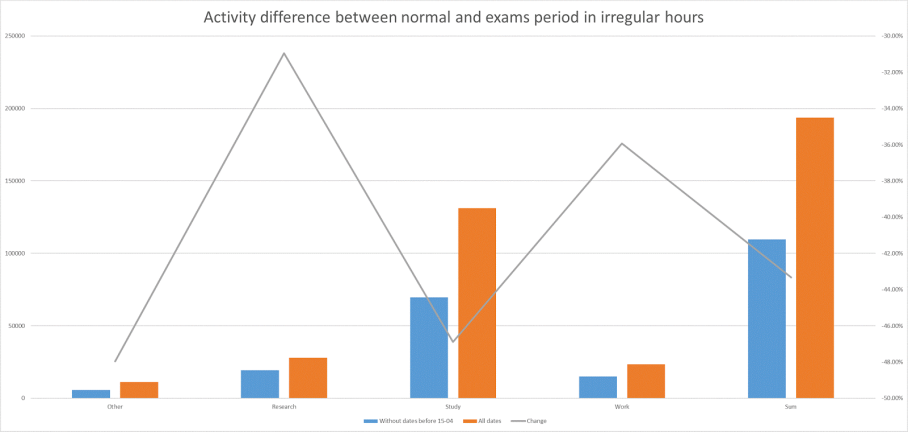


Figure 3.17: Activity difference between normal and exams period in irregular hours

3.6. Quality Control

During each step in the processing, analysis, and visualisation of the data, the quality of each operation, and the validity of each result are assessed. In subsections 3.6.1 and 3.6.2, the validation process is described, the overall accuracy and methods to maintain accurate results are discussed.

3.6.1. Accuracy of occupation classification

This subsection will assess how accurate the classification method was executed, it will compare the numbers produced with numbers provided by FMRE.

During pre-processing there were 46.067 distinct users connected to eduroam, while, according to TU Delft reports there are 30.579 distinct users, without taking into consideration the guests. The 15.000 user difference is, as it is discussed, probably due to the large number of visitors that TU Delft has, users that sometimes get more than one username (around 3000) and some machines that might have not been removed. However, the guests should be regarded as the main reason of the difference. Furthermore, over the total users recognised in the dataset there were 12% more students identified, 7% more academic staff and 4% less support staff, than in TU Delft statistical data. This leads to a Root Mean Square Error (RMSE) of 5% (1528 users), which can be regarded as a very good result. However, it is not known if the users that are counted per occupation are the actual users of this group. Therefore further accuracy measurements are given in Table 3.17.

Table 3.17: Profiling accuracy over total dataset users

	Student	Academic staff	Support staff	other	sum (w/o other)
Calculated data	23582	6815	3010	12660	33407
Reference Data	20980	6349	3128	unknown	30579
Difference	466	2602	-118	-	2950

In the Tables 3.18 and 3.19, the results of the Markov model are illustrated. The students are recognised with 64% accuracy while all other classes are at 0%, for an overall 50% accuracy. Since the ground truth (vertical columns) only consists out of fourteen users these accuracies can not be regarded as significant.

Ov. Accuracy = (7+0+0)/14=50%
Ov. Misclassification = (4+2+1)/14=50%

Table 3.18: Profiling results in known users

ID	Real Occupation	Assigned Occupation
USER 1	Student	Other
USER 2	Student	Student
USER 3	Student	Student
USER 4	Student	Student
USER 5	Student	Academic staff
USER 6	Student	Academic staff
USER 7	Student	Student
USER 8	Support staff	Student
USER 9	Support staff	Academic staff
USER 10	Academic staff	Student
USER 11	Student	Student
USER 12	Student	Academic staff
USER 13	Student	Student
USER 14	Student	Student

Table 3.19: Confusion matrix of classification

	Student	Academic staff	Support staff	Other	Classification overall	Producer Accuracy (Precision)
Student	7	1	0	1	9	77.78%
Academic staff	3	0	1	0	4	0%
Support staff	0	0	0	0	0	No data
Other	1	0	0	0	1	0%
Truth overall	11	1	1	1	14	
User Accuracy	63.63%	0%	0%	0%		

3.6.2. Accuracy of main building

For determining the activity of each user its main building, its occupation and its scanned location are taken into consideration. However, the main building of each user is determined as the building where that particular user spent his most time at. Therefore, this assumption leads to users that have as main building common used buildings that can't match their occupation, for example students with main building library. The aforementioned example, might make sense that a student spends most of his time in library studying, yet to analyse the the accuracy of the main building classification, it needs to be determine to which faculty a user belongs to.

To make an accurate statement independent of the occupation classification of a user the total numbers of faculty staff and students are compared to the numbers of our dataset. Since official statistics do not contain guests of the university they have to be filtered out first. This is done if a user has less than four connection days. With this 13200 users are classified as guests. After this the main building of a user is determined, as his main faculty, but only if he spent more than one hour there. Table 3.20 gives an impression of the accuracy of the home faculty or workplace determination.

Table 3.20: Accuracy of main building determination

	3mE	Arch	CEM	EEMCS	IDE	AE	TPM	AS	Other	Total
Calculated data	4749	4435	4903	3693	3230	2743	1764	3613	1905	31035
Reference data	5346	3698	4610	3529	2435	3163	1877	3840	2099	30597
Difference	-597	737	293	164	795	-420	-113	-227	-194	438

The biggest difference can be found at the industrial design and architecture faculty. The overall count of people per faculty has a RMSE of 1.5% (462 users) and can therefore be considered as good.

4

Cross-cutting themes

4.1. Validity and Accuracy

Due to the design of the tracking infrastructure, a user has a minimum connection time of five minutes. Hence, the precision of the session duration is five minutes. If a user is only passing by an AP and gets picked up, he will still have a connection of at least five minutes. However, if a user has multiple consecutive sessions with duration of five minutes, those sessions are aggregated into one longer session.

Due to the nature of Wi-Fi tracking, a user detected in a certain AP is not necessary located around that AP. Meaning that, a user might be passing outside a building and he might be scanned by an AP inside it, but actually he is not there. To determine the accuracy of scanned users one has to test each AP to determine the overall accuracy. In this project there are around 1700 APs involved, thus it is difficult to determine the accuracy of each one of them. However, given the 5 minute minimum duration rule that is implemented above, people that are just passing by, should be able to be removed.

According to the previous observations, false location of scanned users can also happen between floors of a building, hence the accuracy of the location determination on AP level is regarded as not good enough. Given that a user moving between floors will not be removed, due to consecutive five minute scans, the AP information is not utilised. Instead, information on floor level is regarded more accurate.

The system is designed in a way that a device can only maintain a single connection at a time. This means if a user changes his location inside the five minute refresh rate of his last access point, a new connection will only be saved after these five minutes. The locations a user visited in between will not be recorded. However, the wifilog contains 10300 records where devices have more than one connection at a time (see subsection 3.2.1 - Invalid session duration). If these connections are not corrected, the total session duration of the corresponding users can be overrepresented by several hours.

4.2. Representativeness

Representativeness of the data collected reflects what user categories the Wi-Fi tracking system implemented can identify. In this section the user categories that the Wi-Fi tracking system cannot record will be presented.

The first big category is the users that were not connected to the eduroam Wi-Fi network. This category consists of smaller categories, a) those who can't access the network because of old technology, b)

those who can't access it because they are not part of the academic network that has access to the eduroam network, c) those who consciously turned off their Wi-Fi devices, d) those who connected to the network through wired connection. The outcomes of section 3.6.1 suggest that the number of these users is not significant and is merely found inside the support staff.

The second category refers to places that there are no available data in the database given. The missing data account for either buildings that are excluded for security reason, e.g. Nuclear reactor, or for places that are not covered by access points, hence no information is collected.

4.3. System of Access Points

The content of this section describes to what degree the system of APs is regarded appropriate for measuring and tracking. Also, the missing and the essential parts to make the system work correctly, are presented below.

First of all, in the dataset that is given, every device is usually detected in only one AP. Thus, with the information given, the processes of fingerprinting and trilateration, that would allow to determine the position of a device, cannot be carried out. Therefore, it is assumed that the information that the Wi-Fi scanners provide are correct and it is assumed that the detected devices are closer to the specific access points that they are detected its different time.

Furthermore, the maps with the exact position of the APs in the campus are not provided. This information is considered essential for having better quality in the project's results. With the map of the APs locations, the use of the areas around each AP can be distinguished in different categories, such as lab, offices, rooms etc. By using it, the activity of the people that are detected in different areas can be classified more accurately as laboratory work, work or study, than through the framing of the profiles to the users that is done in this project.

5

Data protection

This project is dealing with big data that is closely linked with users of the TU Delft network. Personal data is any information relating to an identified or identifiable natural person (European Parliament and European Council, 1995), therefore it is very important to take into account the Data Protection Directive (DPD) or called Directive 95/46/EC. This directive needs to be taken into consideration in this project, in order to provide a legal result.
The DPD is an European Union directive adopted in 1995 which regulates the processing of personal data within the European Union (EU). The DPD is an important component of the EU privacy and human rights laws. In The Netherlands this directive is used as the basis for the Dutch law on the protection of personal data. In April 2016, the European Council and Parliament adopted the General Data Protection Regulation (GDPR). The GDPR replaced the DPD as this one is outdated due to the digitalized world since 1995. To allow e.g. companies to adjust to the GDPR there is a two-year transitional period after which they can be fined when not acting according to the GDPR. The difference with a directive is that the GDPR does not require any enabling legislation to be passed by the member state governments. For this project, the DPD is used as the legal framework, this means that if this project is redone in the future, the new GDPR needs to be considered as the legal framework (European Parliament and European Council, 2016).
In section 5.1, a general description of the dataset (wifilog) that is provided for the specific project, is analysed. The legitimate grounds, the principles and the data subject rights are presented in section 5.2, 5.3 and 5.4, respectively. Some general remarks on the attitude of TRACK-id, as the data processors, are mentioned in section 5.5.

5.1. Project dataset

The TU Delft is considering to make the wifilog table open data in the near future. However, the wifilog in the current state can be regarded as personal data. The username and MAC address of individuals are hashed, yet with each new update of the wifilog, the username and MAC address are hashed in the same way. Beside the username and MAC address, the code of the AP (apname) to which the connection was made, is saved. This apname describes the location of the individual quite accurately, due to the limited range of the eduroam APs and the high density of the APs on the TU Delft campus. At the moment, there are no maps available with the locations of the APs, nonetheless people could create their own map as the code of the AP is printed on the AP. If this information is combined with some information about an individual, then it is possible to identify him or her in the wifilog. To give an example: it is known that a certain individual arrived at 06-06-2016 09:00 at AP A-08-B-102 (AP located in the Faculty of Architecture and the Built Environment) and used his or her smartphone for an hour and then left. With the additional information that this individual normally studies at the Faculty of Applied Sciences, it is very easy to identify this person in the wifilog.

Even if the column username would be removed from the wifilog, it is still regarded as personal data. It becomes a little harder to link all the devices to a unique user. Notwithstanding, the username is not needed, in order to identify a person based on his or her MAC address. If the column MAC address would be removed, the column username makes the wifilog definitely personal data. In case both columns would be removed and only the locations of the users are stored, the data cannot be easily used to answer the questions of FMRE, because it is impossible to determine the amount of users. In addition, if this data is combined with e.g. camera data, it will still be possible to count the amount of users and be able to identify persons. Hence, even if the columns username and MAC address are removed, the wifilog can still be regarded as personal data and used accordingly to the DPD.

5.2. Legitimate Grounds

According to the DPD, collecting and processing of personal data of individuals is only legitimate in one of the following circumstances (Article 7 of the (European Parliament and European Council, 1995)), (European Commission, 2016):

- the data subject, which is the individual, has unambiguously given his or her consent after being adequately informed; or
- the processing is needed for the performance of a contract to which the data subject is party or in order to take steps at the request of the individual prior to entering a contract; or
- processing is required for compliance with a legal obligation to which the controller is subject; or
- processing is necessary in order to protect the vital interests of the individual; or
- processing is necessary to perform tasks of public interests or tasks carried out by the government, tax authorities, the police or other public bodies; or
- the data controller or a third party has a legitimate interest in doing so, as long as this interest does not affect the interests of the data subject, or infringe upon his or her fundamental rights, in particular the right to privacy. This establishes the need to find a reasonable balance between the data controllers' business interests and the privacy of the data subjects.

For the GSP the teams of students have to address the questions of the stakeholder FMRE. In order to answer the questions of FMRE the ICT department of the TU Delft provides the needed data. The ICT departments gathers and stores the data and is therefore the owner of the dataset (wifilog). This makes the TU Delft being the data controller and TRACK-id the data processor, in order to answer the questions of FMRE.

5.3. Principles

Moreover, when processing the personal data TRACK-id has to respect several principles relating to the quality of the data (European Parliament and European Council, 1995), (European Commission, 2015):

- (a) personal data should be processed fairly and lawfully;
- (b) data must be collected for a specific, explicit and legitimate purpose and not further be processed in a way incompatible with those purposes;
- (c) the data must be adequate and not excessive in relation to the purpose for which it is collected and/or further processed;
- (d) the data must be accurate and update when necessary; every reasonable step must be taken to ensure that data which are inaccurate or incomplete are erased or rectified;

- (e) personal data must not be kept any longer than strictly necessary for the purposes for which the data were collected or for which they are further processed. Data controllers must protect personal data against accidental or unlawful destruction, loss alteration and disclosure, particularly when processing involves data transmission over networks. They shall implement the appropriate security measures. These protection measures must ensure a level of protection appropriate to the data.

The controller of the personal data is responsible that the data is used, processed, and stored in such a way that it is compliant with the principles specified above (Article 6 (2) of (European Parliament and European Council, 1995)). According to points b, c and e, the data should be collected, stored and processed only for as long as the purpose of the project demands. The project's scope of research is the determination of the amount of users and their activities outside of opening hours, in TU Delft buildings.

5.4. Data Subject Rights

The data controller, in this occasion TU Delft, is recommended to take into consideration the information that needs to be provided to the data subject and the data subjects rights. According to the Directive 95/46/EC, the data subject has the right to access the data and to object. The data subject can obtain the following information from the controller (Articles 11, 12 and 14 of the Directive 95/46/EC):

INFORMATION TO BE GIVEN TO THE DATA SUBJECT

- the identity of the controller and of his representative, if any;
- the purposes of the processing for which the data are intended;
- any further information such as
  - the recipients or categories of recipients of the data,
  - whether replies to the questions are obligatory or voluntary, as well as the possible consequences of failure to reply,
  - the existence of the right of access to and the right to rectify the data concerning him

THE DATA SUBJECT'S RIGHT OF ACCESS TO DATA

- without constraint at reasonable intervals and without excessive delay or expense:
  - confirmation as to whether or not data relating to him are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed
  - communication to him in an intelligible form of the data undergoing processing and of any available information as to their source
  - knowledge of the logic involved in any automatic processing of data concerning him at least in the case of the automated decisions.
- as appropriate the rectification, erasure or blocking of data
- notification to third parties to whom the data have been disclosed of any rectification, erasure or blocking carried out in compliance with (b), unless this proves impossible or involves a disproportionate effort.

THE DATA SUBJECT'S RIGHT TO OBJECT

- (a) Where there is a justified objection, the processing instigated by the controller may no longer involve those data;
- (b) to object, on request and free of charge, to the processing of personal data relating to him which the controller anticipates being processed for the purposes of direct marketing, or to be informed before personal data are disclosed for the first time to third parties or used on their behalf for the purposes of direct marketing, and to be expressly offered the right to object free of charge to such disclosures or uses.

Therefore, the TU Delft as a data controller should inform the data subject about the project's subject and needs. If individuals don't want to be in the dataset provided to TRACK-id, they can send a request to TRACK-id to opt-out, so they will be removed from the dataset that TRACK-id uses, in order to answer the questions of FMRE.
FMRE has a legitimate interest in asking the questions, because they will use the results of this research to gain knowledge in the usage of the TU Delft campus and how to adjust the TU Delft campus in the future.

5.5. Data processors

Considering the above, TRACK-id's purpose, according to the data controller's needs, is clear and the security of the data is taken into consideration. TRACK-id is ensuring the security of the wifilog dataset using a password secured database during the process and the deletion of it after the completeness of the project. Furthermore, the outcomes of this project do not include any information about individuals, but only aggregated data. TRACK-id is assigning profiles to the users, such as student, academic staff, support staff or other based on general assumptions, in order to answer the question of FMRE. The framing of profiles is not covered in the current DPD, however in the next version from the 25th of May 2018, onwards profiling will be included. Until then, it is regarded acceptable to profile people according to a framework, but it is not acceptable to act upon this profiling. During the profiling it is important to protect the sensitive data of the data subjects, thus it is important not to collect or process data that fall into sensitive categories of data, such as public health and social protection. Moreover, it is important to mention that scientific research has to be respected.

6

Discussion and future work

In this section the problems that have been detected during this project regarding the Wi-Fi network and some problems that occurred at the methodology, are explained.

6.1. Identified Wi-Fi data problems

It is important to mention that the outcomes of this project cannot comply completely with the reality, considering that not all users of a building have access to the Wi-Fi network all the time that they actually are on the campus. Also, some people might never use the Wi-Fi network. These are sessions that cannot be found through the dataset and cannot be solved in this project.

A user can have more than one device, which means that at the same time some sessions of a user from two or more different devices are overlapping. This problem is solved in subsection 3.2.3, with the aggregation of the time from the different devices of a user.

In subsection 3.2.2, the splitting at midnight is done when a user stays after 00:00:00 o'clock. In this case, his session is split into two different days and the aggregating time is not considered as one whole session, as it is in reality, but in two different sessions, one for each day.

Furthermore, some invalid records (see subsection 3.2.1), are detected through the dataset. Two records of the same user and the same device, might overlap. Usually, one of the two records continues for many hours. The big session is usually wrong and such an example is detected and examined in the dataset, for one of the known users.

Access to the wireless network of other universities and HBO organisations (universities of professional education) in the Netherlands, is also possible via eduroam. For this reason, it is possible to have some guest users in the dataset of the TU Delft campus. These users are taken into consideration in the framing of profiles to the users in subsection 3.3.1/3.3.2. Moreover, this use of the eduroam has increased the usernames in the dataset. Also, it is very common for an international student to have an eduroam password from another university. In this case, that user has more usernames and some problems on the results might derive. A problem that derived during the exploitation of the dataset is that some devices are connected to more than one username. It is found that a specific username appears in many different devices of different users. This username is the anonymous and it is solved in subsection 3.2.2. In order to connect with eduroam, the connection can be done with TTLS or with Protected Extensible Authentication Protocol (PEAP). To connect with the wireless network eduroam via PEAP the Operating System has to support PEAP or needs a supplicant which is PEAP supported. PEAP is integrated in the Operating System of the latest versions of Windows.

To connect with eduroam via TTLS (Tunneled Transport Layer Security) the 802.1x supplicant is required. The 802.1x security standard is used for authentication and an 802.1x supplicant is therefore required. (TU Delft, 2016b)

Sometimes the users might log in from a different device. These devices appear in the dataset with the username that logged in the specific time. Therefore, a username might be connected with more devices than the number of devices that he actually has. In order to solve this problem, at the preprocessing of the dataset, the users with more than the expected number of devices are not being removed as outliers. (TU Delft, 2016c)

The first record of a user is automatically assigned as a five minutes session duration. This cannot always be correct.

Some usernames are related to print machines etc., these devices provide wrong results on the amount of the users during the regular and the irregular hours. This is not solved. The coverage of the Wi-Fi network might not be good for all the campus. Its limitations need to be investigated for future improvements on the quality of the results.

6.2. Improvements upon the applied methodology

The exceptions of the opening hours, for example the extension of the opening hours of the library during the exam periods have not been taken into consideration on the definition of the regular_hour column in wt3_wifilog table.

1. Quality assessment of the number of the users during the irregular hours should be performed.
2. For the profile of the users ,all buildings should have been used, even the ones that are out of the scope of our research. The results of the profiles would have been more accurate.
3. The questionnaires have not been utilised, hence the training sets are based on rough assumptions.
4. User occupation of a user group in one time slot (one hour) at a specific map location or access point can be implemented in future development.
5. Time usage in different buildings of a single user. Due to privacy issues, the username is hashed and is not connected to the actual users. In further work, it is might be feasible that a password is set for the application by the FMRE department (who knows the real NetID) to limit access; or it is also reasonable for user to login with NetID and password and track only him/herself.

7

Conclusion

The scope of this research is the determination of the amount of users and activities outside opening hours. The project has succeeded up to a point to deliver the required results.

Initially, the preprocessing recognised 46.067 distinct users that were connected to eduroam, while, according to TU Delft reports there are 30.579 distinct users, without taking into consideration the guests. The 15.000 user difference is, as it is discussed, probably due to the large number of visitors that TU Delft has, an anomaly where users sometimes get more than one username and some machines that might have not been removed. However, the guests should be regarded as the main reason of the difference. Most of the were successfully classified into the group Others.

Secondly, the users could be accurately assigned to a faculty or main building. All the faculties are represented well inside the Wi-Fi data. Additionally, this leads to a good foundation for the activity assignment, which is partially based on the usual location of a user.

Thirdly, during irregular hours there were more than 100.000 accumulate users. Almost 70% of this were students. It is regarded that the student number is high, since the exceptions in buildings were not taken into consideration while counting users during irregular hours. Given the fact that the first two weeks were exam period in TU Delft and some buildings stay open for extended periods, there were many students that were counted to be there as in irregular hours, while actually the buildings were still open. Additionally, the approximate time that activities took place during irregular hours is around 130.000 hours of study, almost 28.000 hours of research, more than 23.000 hours of work and 11.000 hours of other activities. Again, the study hours are overestimated, because of the false exceptions. If the days until the end of exams are not taken into consideration, 24.6% of days (15 out 61 days) are removed and each occupation category during irregular hours drops around 36%, which is the first indication implying that those days had more users than usual during irregular hours. Moreover, by dropping those days, research and work activities drop by 31% and 36% accordingly, while study and other activities drop by 47% and 48%. In other words, almost 25% of days hold close to 50% hours spent on those activities during irregular hours. However, in case the exception are taken into consideration the results should get better.

Finally, regarding the data protection, since the project, the process and the data have been handed to the TRACK-id team by TU Delft, TRACK-id is regarded as the data processors and TU Delft as data controller of this project. The TRACK-id team kept the data safe in a password protected database during the project, did not track sensitive data, aggregated the data to ensure anonymity and took no direct act using the created profiles. Also, it will ensure that the data will be deleted after the end of the project. On the other hand, TU Delft, as data controller should make sure that the data is collected, stored and processed only for the duration of the project. It should respect the data subject's rights to access the data and to object and also it should offer all the information that are required, according to

section 5.4. Lastly, in this project the Data Protection Directive (European Parliament and European Council, 1995) is utilised since it is still active, even if it has been replaced by General Data Protection Regulation in April 2016. However, should TU Delft consider publishing eduroam as public data, the newest law should be taken into consideration.

8

Recommendations

- The preprocessing can be improved to further identify and handle machine users.
- The Wi-Fi tracking infrastructure can be improved to allow more tracking techniques, however the TRACK-id project team strongly recommends against it since tracking with the current infrastructure and available techniques is already regarded as intrusive tracking technique.
- The Wi-Fi tracking accuracy could be improved and validated in order to have accurate results for the connected users.
- The exception in closing hours of buildings should be taken into consideration for more accurate count of users during irregular hours
- The outcomes from the model are not considered accurate enough for immediate use. In order to increase the model's accuracy it is regarded necessary to improve the training sets of the profiles. In order to improve the training sets it is proposed to use the questionnaires.
- The data protection issues that are mentioned in chapter 5, are very important. It is recommended that the TU Delft as data controllers of this project, should consider the general remarks that are presented above. Some of the basic recommendations are the notification to the users of the TU Delft network about the purpose and the aim of this project. Moreover, the possibility of opting-out and the possibility of retrieving their personal information through the dataset. TU Delft should balance their interests and the privacy of the data subjects.

Bibliography

N. Abedi, A. Bhaskar, E. Chung, and M. Miska. Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses. *Transportation Research Part C: Emerging Technologies*, 60(November 2015):124–141, 2015. URL http://www.sciencedirect.com/science/article/pii/S0968090X15003113.

D.R. Anderson and K.P. Burnham. Avoiding Pitfalls When Using Information-Theoretic Methods. *The Journal of Wildlife Management*, 66(3):912–918, 2002.

G.E.P. Box and N.R. Draper. *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, 1987. URL http://tocs.ulb.tu-darmstadt.de/205603718.pdf.

G.E.P. Box, J.S. Hunter, and W.G. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery*. Wiley, 2nd edition edition, 2005. ISBN 978-0-471-71813-0.

CBP. Wifi-tracking van mobiele apparaten in en rond winkels door Bluetrace. Technical report, College bescherming persoonsgegevens, 2015. URL https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/rapport_db_bluetrace.pdf.

K. Christensen, R. Melfi, B. Nordman, B. Rosenblum, and R. Viera. Using existing network infrastructure to estimate building occupancy and control plugged-in devices in user workspaces. *International Journal of Communication Networks and Distributed Systems*, 12(1):4–29, 2014.

European Commission. Obligations of data controllers, 2015. URL http://ec.europa.eu/justice/data-protection/data-collection/obligations/index_en.htm. (Accessed: 4 June 2016).

European Commission. Collecting & processing personal data: what is legal?, 2016. URL http://ec.europa.eu/justice/data-protection/data-collection/legal/index_en.htm. (Accessed: 4 June 2016).

European Parliament and European Council. Directive 95/45/ec. *Official Journal L*, 281(23/11/1995): 0031–0050, 1995. URL http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML.

European Parliament and European Council. Regulation (eu) 2016/679. *Official Journal of the European Union*, L 119(1), 2016. URL http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

M. Harwood. *CompTIA Network+ N10-004 Exam Cram*, chapter Wireless Networking. Pearson IT Certification, Part of the Exam Cram series, 3rd edition edition, 2009. URL http://www.pearsonitcertification.com/articles/article.aspx?p=1329709&seqNum=3.

O.D. Lara and M.A. Labrador. A Survey on Human Activity Recognition Using Wearable Sensors. *IEEE Communications Surveys & Tutorials*, 15(3):1192–1209, 2012.

M.J.P.M. Lemmens. GIS: THE DATA PROBLEM. In *Conference on Geographical Information Systems*, 1990.

M. Lin, H. Cao, V. Zheng, K.C.C. Chang, and S. Krishnaswamy. Mobility Profiling for User Verification with Anonymized Location Data. *IJCAI International Joint Conference on Artificial Intelligence*, 2015 (January):960–966, 2015.

S. Lühr, H.H. Bui, S. Venkatesh, and G.A.W. West. Recognition of human activity through hierarchical stochastic learning. IEEE, 2003.

R. Mautz. *Indoor Positioning Technologies*. ETH Zurich, Department of Civil, Environmental and Geomatic Engineering, Institute of Geodesy and Photogrammetry, 2012. doi: http://dx.doi.org/10.3929/ethz-a-007313554.

M. Mühlenbrock, O. Brdiczka, D. Snowdon, J.L. Meunier, et al. Learning to detect user activity and availability from a variety of sensor data. IEEE, 2004.

Valery A Petrushin. Hidden markov models: Fundamentals and applications. In *Online Symposium for Electronics Engineer*, 2000.

M. Stamp. A revealing introduction to hidden markov models, 2015.

TU Delft. Campus Development, 2013a. URL https://intranet.tudelft.nl/en/on-campus/faculties-and-services/university-corporate-office/fmre/campus-development/. (Accessed: 1 April 2016).

TU Delft. Campusvision TU Delft, 2013b. URL https://issuu.com/tudelft-mediasolutions/docs/campusvision_a5_def_-_the_vision_fo. (Accessed: 1 April 2016).

TU Delft. Housing TU Delft, 2016a. URL http://www.tudelft.nl/en/about-tu-delft/contact-and-accessibility/housing-tu-delft/accessibility/. (Accessed: 1 April 2016).

TU Delft. Connecting with eduroam via peap, 2016b. URL https://intranet.tudelft.nl/en/services/fmreict-psc/network/wireless-network/eduroam-wireless-network/connecting-with-eduroam-via-peap/. (Accessed: 16 June 2016).

TU Delft. Connecting with eduroam with ttls, 2016c. URL https://intranet.tudelft.nl/en/services/fmreict-psc/network/wireless-network/eduroam-wireless-network/connecting-with-eduroam-with-ttls. (Accessed: 16 June 2016).

University of Waterloo Campus Community. Factors affecting Wi-Fi service, 2013. URL https://uwaterloo.ca/information-systems-technology/services/eduroam/campus-wi-fi-infrastructure-and-quality-assurance/factors-affecting-wi-fi-service. (Accessed: 8 June 2016).

Daniël Verlaan. CBP: winkels mogen klanten niet zomaar via wifi volgen, 2015. URL http://www.rtlz.nl/tech/cbp-winkels-mogen-klanten-niet-zomaar-wifi-volgen. (Accessed: 14 June 2016).

Wikipedia. Hidden markov model, 2012. URL https://en.wikipedia.org/wiki/Hidden_Markov_model#/media/File:HiddenMarkovModel.svg. (Accessed 9 June 2016).

A

Outcomes for Facility Management and Real Estate

This appendix contains all the charts created for the buildings of the TU Delft campus that were in the scope of the research. The left page first displays some general information about the building, after which the graphs are explained and some conclusions are made. At the bottom of the left page there is a Figure showing the information about the different user of a certain building and what the user his/her home building is (except for the TU Delft campus). The right page shows two charts, one displaying the amount of user divided by user category and a second chart that shows the amount of hours the different user categories spent at the considered building.

A.1. TU Delft Campus

Type of buildings: educational, facility management and real estate, and research buildings
Considered buildings: 3 5 8 12 15 20 21 22 23 30 31 32 34 35 36 37 38 43 45 46 60 61 62 63 64 66
Opening hours: depending on the specific buildings
Exceptions: none

The Figures A.1 and A.2 show respectively the amount of users and hours spent during the regular and irregulars hours at the TU Delft Campus from the 1st of April until the 31th of May. The largest groups of users of the TU Delft Campus in descending order are: students, academic staff, support staff, and others. These users together spent the most time at the following activities in descending order: study, work, other, and the least hours are spent at research.

As can be seen in Figure A.1 the amount of users during the regular hours in the weekends is much smaller compared to the amount of users during regular hours of the weekdays. This is also reflected in the amount of hours spent (Figure A.2). However, for the irregular hours the amount of hours spent is commonly higher during the weekends than the amount of hours spent during the irregular hours during the weekdays. For the irregular hours the differences are not so big for the amount of users or are sometimes even comparable for the weekend days and the weekdays, for the hours spent during the irregular hours the number of hours spent is varying a lot throughout the two month period.

One thing that can be noticed from both the Figures A.1 and A.2 is that the amount of users and hours spent on the 15th of April, the 18th of April, the 27th of April, the 5th of May, the 6th of May, and the 16th of May are much lower compared to comparable weekdays. The large amount of users during the irregular hours from the 1st of April until the 14th of April are due to the extended opening hours of some buildings at the TU Delft Campus due to the exam period. The smaller amount of users on the 15th of April is due to the fact that it was the last day of the exam period. Monday the 18th of April was the first day of the new quarter, not all courses started that Monday, which explains why the number of users is lower at the TU Delft Campus. The 27th of April, the 5th of May, the 6th of May, and the 16th of May are national Dutch holidays and most buildings of the campus are actually closed, which means that the number of users and the hours spent during the regular hours are probably not the correct numbers.

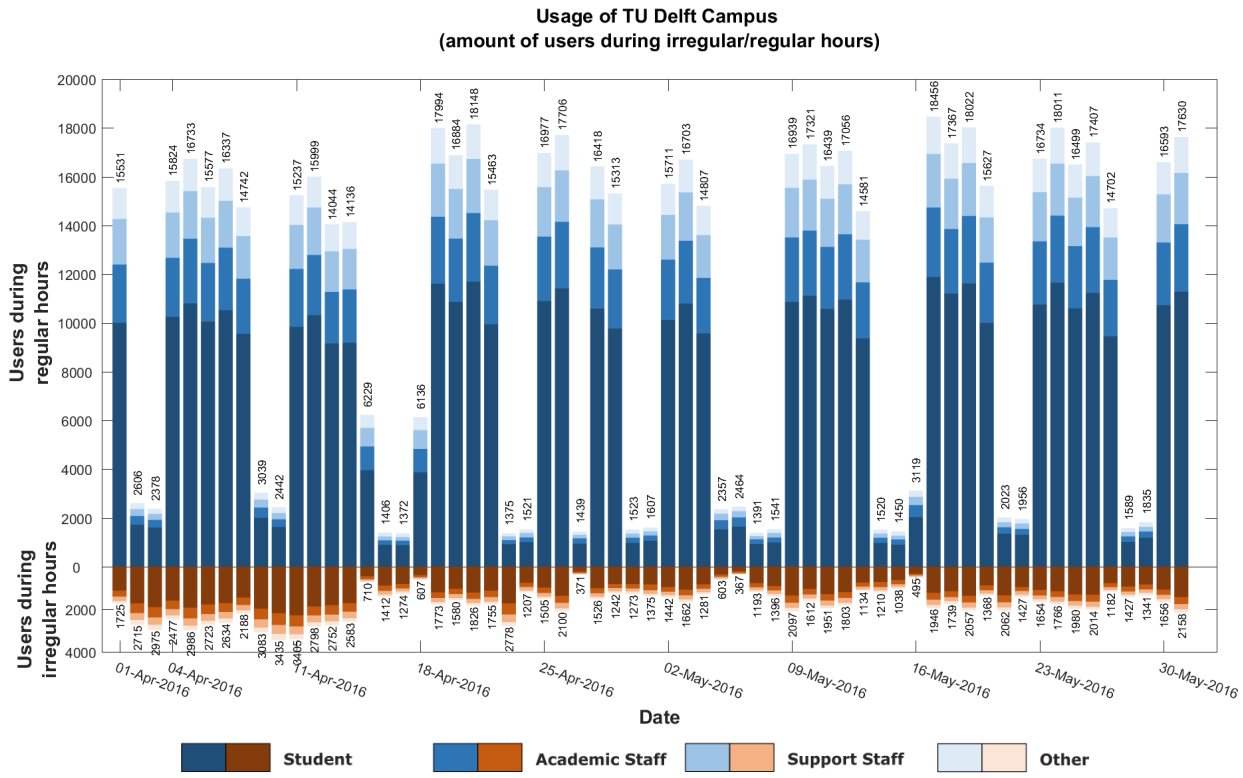


Figure A.1: Usage of TU Delft Campus, amount of users present, divided by type of user

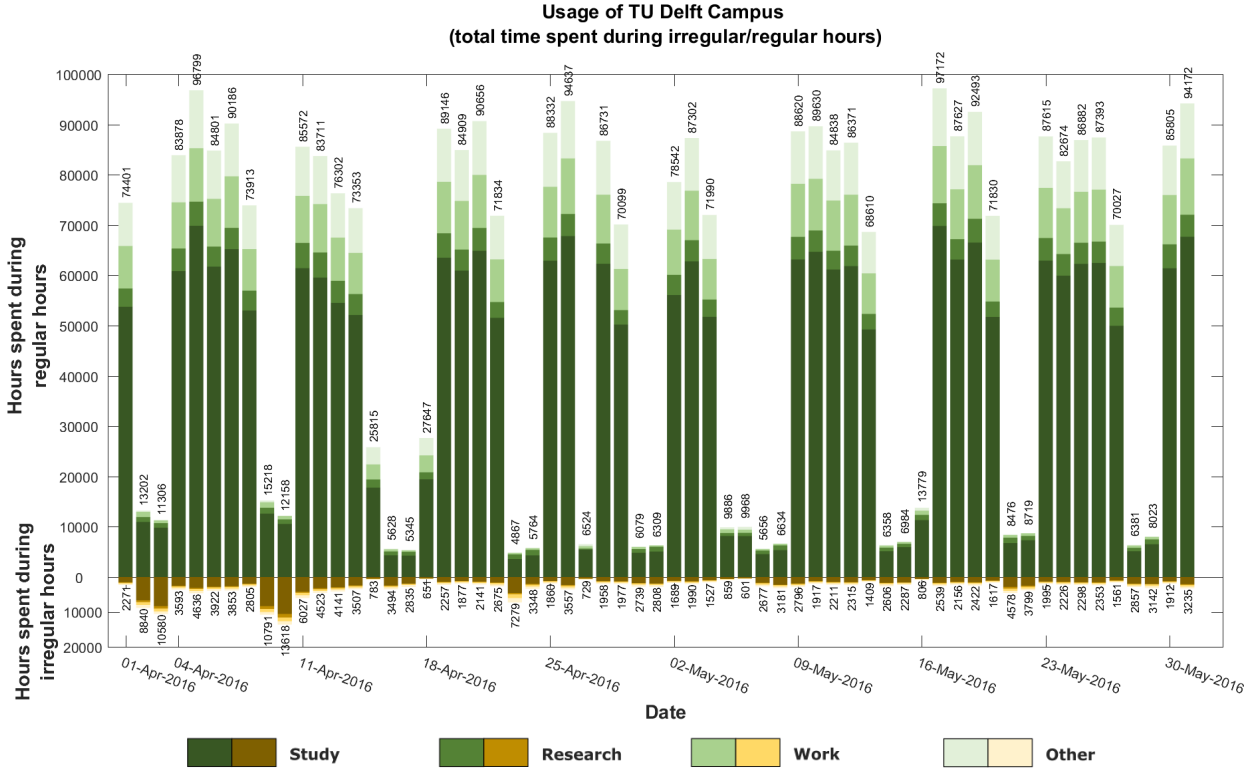


Figure A.2: Usage of TU Delft Campus, the total time spent by users divided into activities

A.2. Building 3: Science Centre Delft

Type of building: educational building
Opening hours: Monday to Friday 8:00 - 20:00
Exceptions: none

Figure A.4 shows that there are many people in the Science Centre Delft on the 20th of April both during the regular and irregular hours. When looking at the total hours spent that day at the Science Centre they are comparable to other days with much less users. This means that the people who were visiting the Science Centre Delft on the 20th of April were staying on average for a short time.

On average there are more people on Saturdays than Sundays at the Science Centre Delft.

The 26th of April they are just a normal number of users at the Science Centre Delft, however they spent more hours than average there.

User group information The building is mainly used by people from itself (in total 1201 hours, on average 20 hours per day by all the users); therein, student (70% of the session time), support staff (16%), academic staff (12.5%) and other (1.7%). The building is also used by people from architecture building (less than 3 hours total session time from all the users per day), TU Delft library (less than 1.5 hours), industrial design engineering (around 1 hour) and 3ME (around 1 hour).

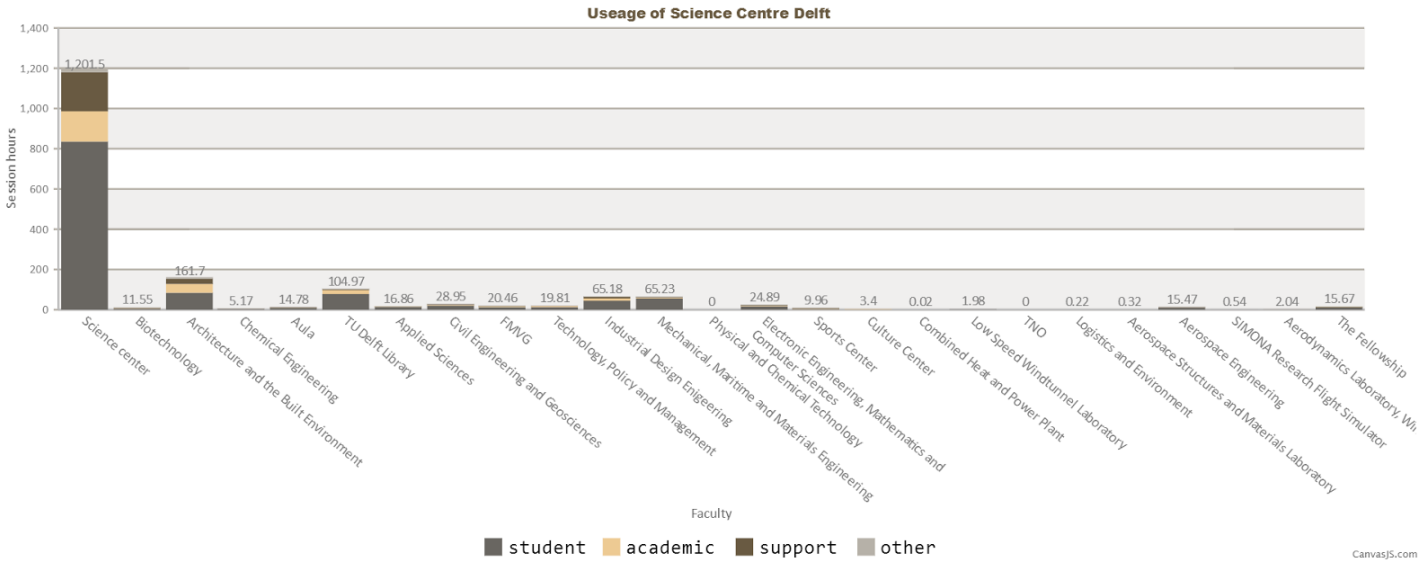


Figure A.3: Information of user group of building 3

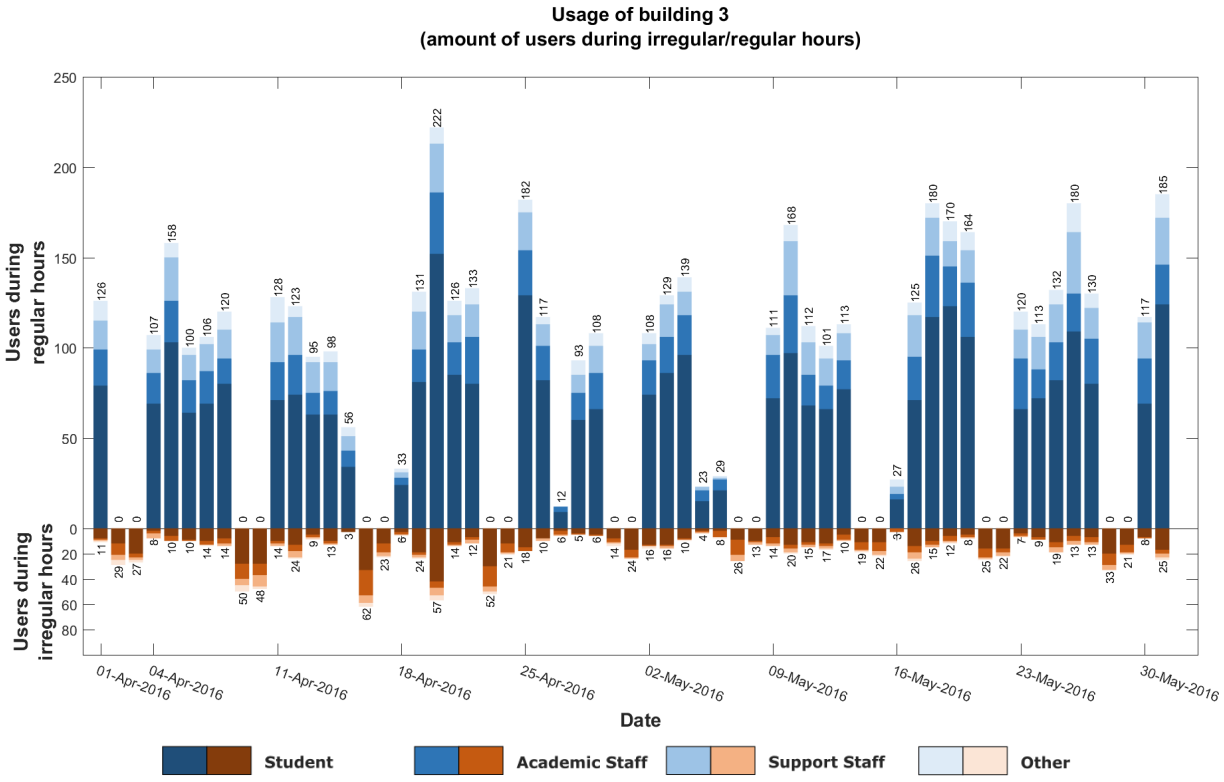


Figure A.4: Usage of building 3, amount of users present, divided by type of user

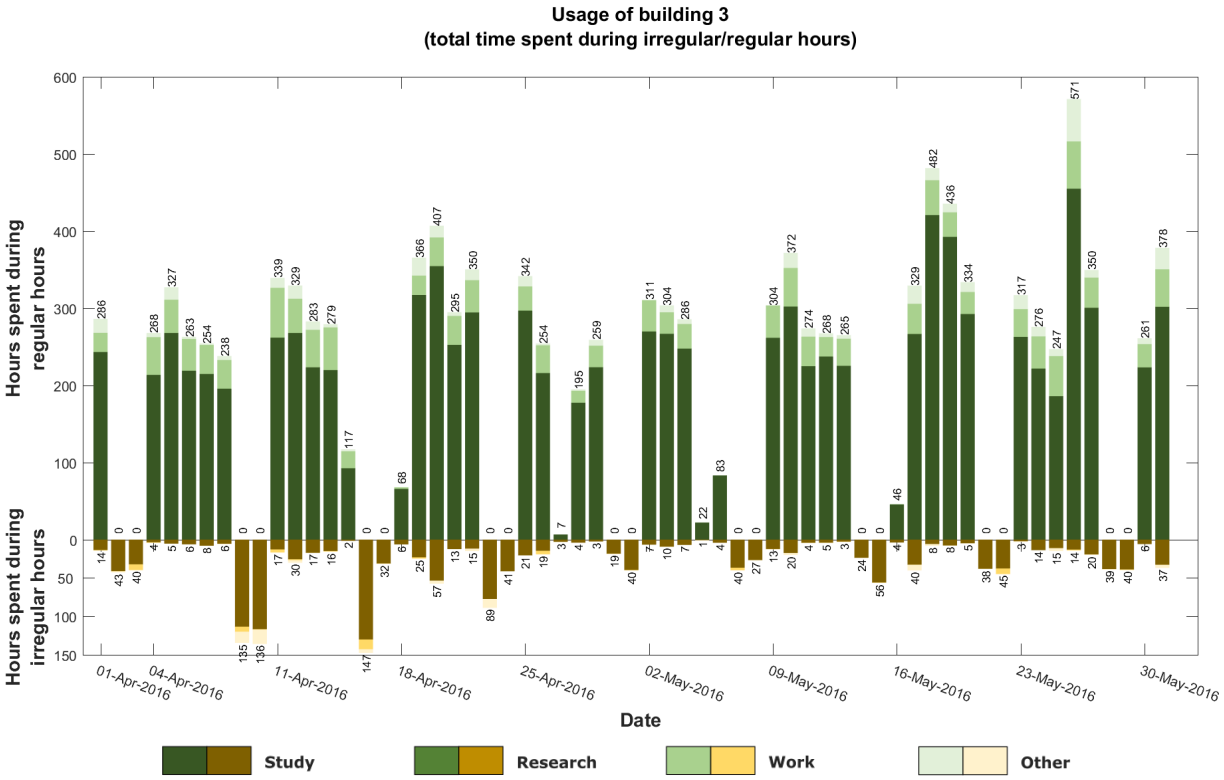


Figure A.5: Usage of building 3, the total time spent by users divided into activities

A.3. Building 5: Department of Biotechnology

Type of building: faculty building
Opening hours: Monday to Friday 07:30 - 18:00
Exceptions: authorized campuscard after closing

User number and time spent during irregular/regular hours
Department of Biotechnology is more use during irregular hours compared with other educational buildings. During the exam week, around 80 people use the building every weekday in irregular hours (for 4.5 hours per person); around 45 people use it every weekend (for 2 hours per person).
After the exam, the building is more used in regular hours and the frequency of being used during irregular hours decreases. The daily user number in weekdays in regular hours is around 280-300 (each person spends about 4 hours per day); the number for irregular hours varies from 38 to 118 but is 50 on average (each person spends 2 hours per day). Main users of the building are students and the most common activity is study.

User group information
The building is mainly used by people from itself (in total 5315 hours, on average 88 hours per day by all the users); therein, student (61% of the session time), academic staff (14%), support staff (8.9%) and other (15.8%). The building is barely used by people from other buildings; people from Chemical Engineering uses the building for around 4 hours (total session time from all the users) per day, people from TU Delft library (less than 3 hours).

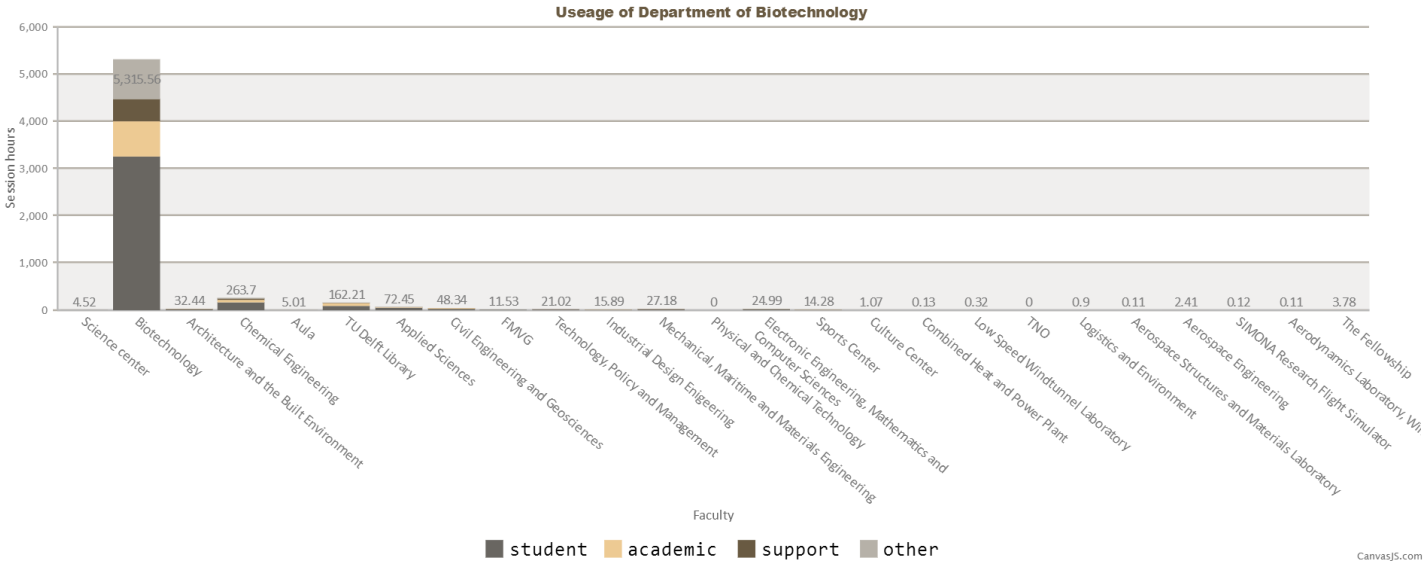


Figure A.6: Information of user group of building 5

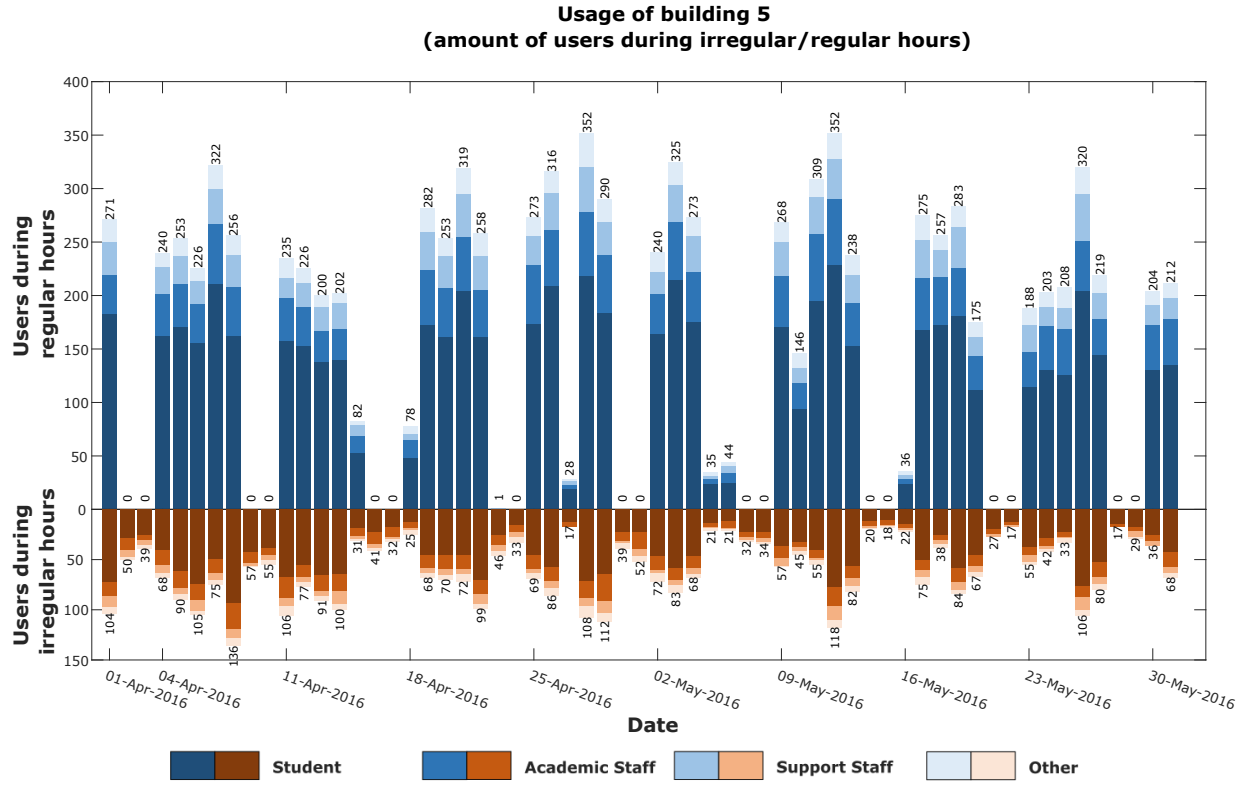


Figure A.7: Usage of building 5, amount of users present, divided by type of user

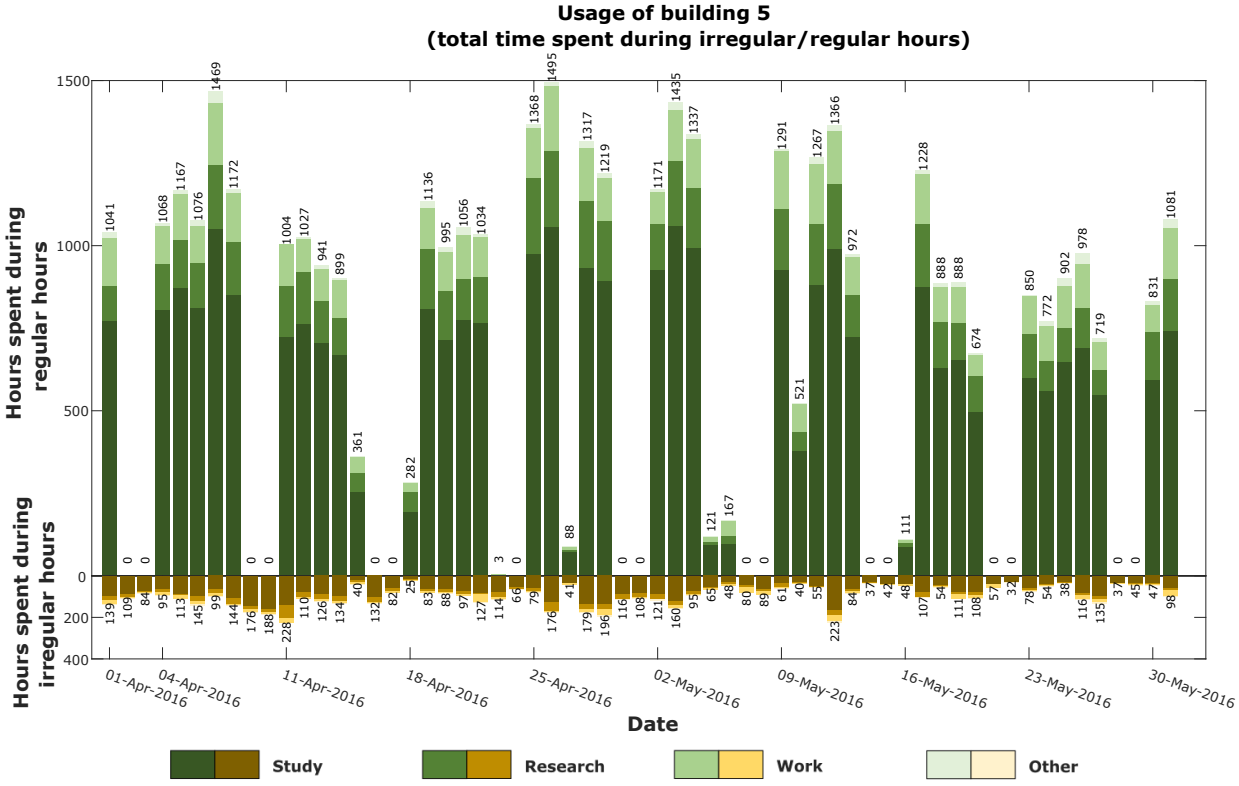


Figure A.8: Usage of building 5, the total time spent by users divided into activities

A.4. Building 8: Faculty of Architecture and the Built Environment

Type of building: faculty building
Opening hours: Monday to Friday 07:00 - 22:00
Exceptions: Friday 07:00 - 19:00

User number and time spent during irregular/regular hours
The amount of users during the irregular hours can be explained with the knowledge TRACK-id has about the behaviour of students at the faculty of Architecture and the Built Environment. The students tend to stay in the building even if the building officially closes at 22:00, the students tend to stay until they get kicked out by the security of the faculty of Architecture and the Built Environment.
The large number of users and the hours they spent at the 2nd of April are due to the fact that the faculty was opened for the parents day organised for the parents of the first year students of the bachelor of architecture.
The 22nd of April the faculty festival of the faculty of Architecture and the Built Environment called BK-Beats was held from the 22nd of April 22:00 until the 23rd of April 04:30. This explains the large amount of users and hours spent during the irregular hours of these days.
On Saturdays and Sundays except for the 2nd of April, and the 23th of April there were between the 53 and 122 users who spent together between the 57 and 135 hours. With the Figures A.10 and A.11 this usage of the faculty of Architecture and the Built Environment cannot be explained. It could be possible that the dashboard which shows the information in these graphs for each hours could provide more insight in what is going on during the weekends at the faculty of Architecture for the Built Environment. The fact that the amount of users during the irregular hours on Friday is always lower compared to other weekdays. This is due to the fact that the faculty closes on Fridays at 19:00 which is not considered in the calculation on the amount of users during the regular/irregular hours. So there are users classified as using the building during the regular hours while they are actually using it during the irregular hours.

User group information
The building is mainly used by people from itself (in total 60,829 hours, on average 1015 hours per days by all the users); therein, student (64%), academic staff (15.74%), support staff (10.5%) and others (9.4%). The building is barely used by people from other buildings; people from TU Delft library use the building for 30 hours per day on average.

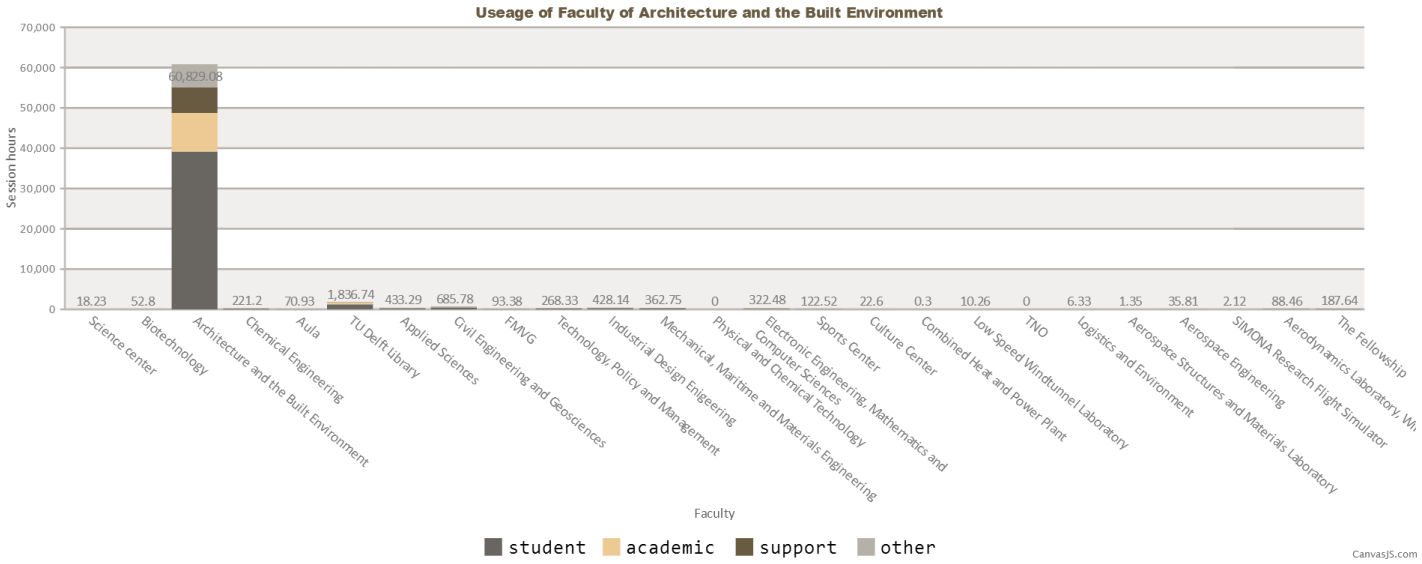


Figure A.9: Information of user group of building 8

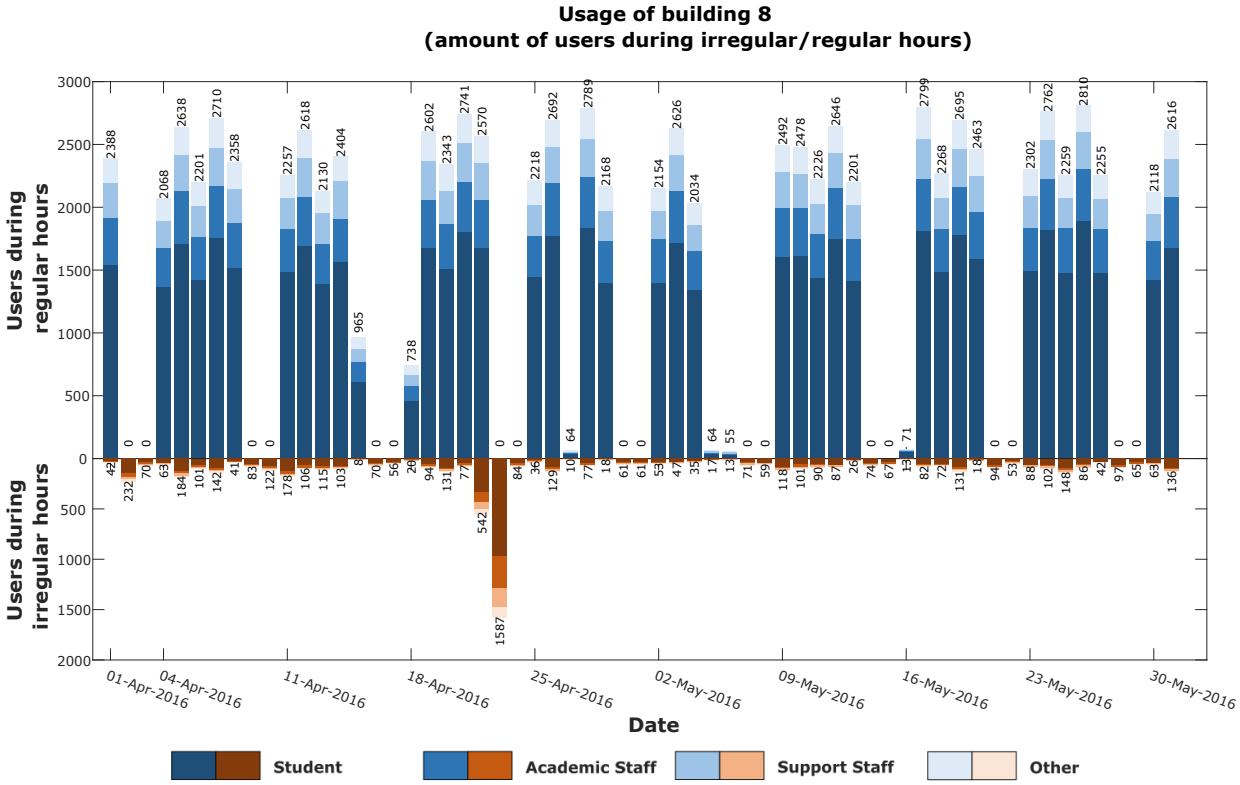


Figure A.10: Usage of building 8, amount of users present, divided by type of user

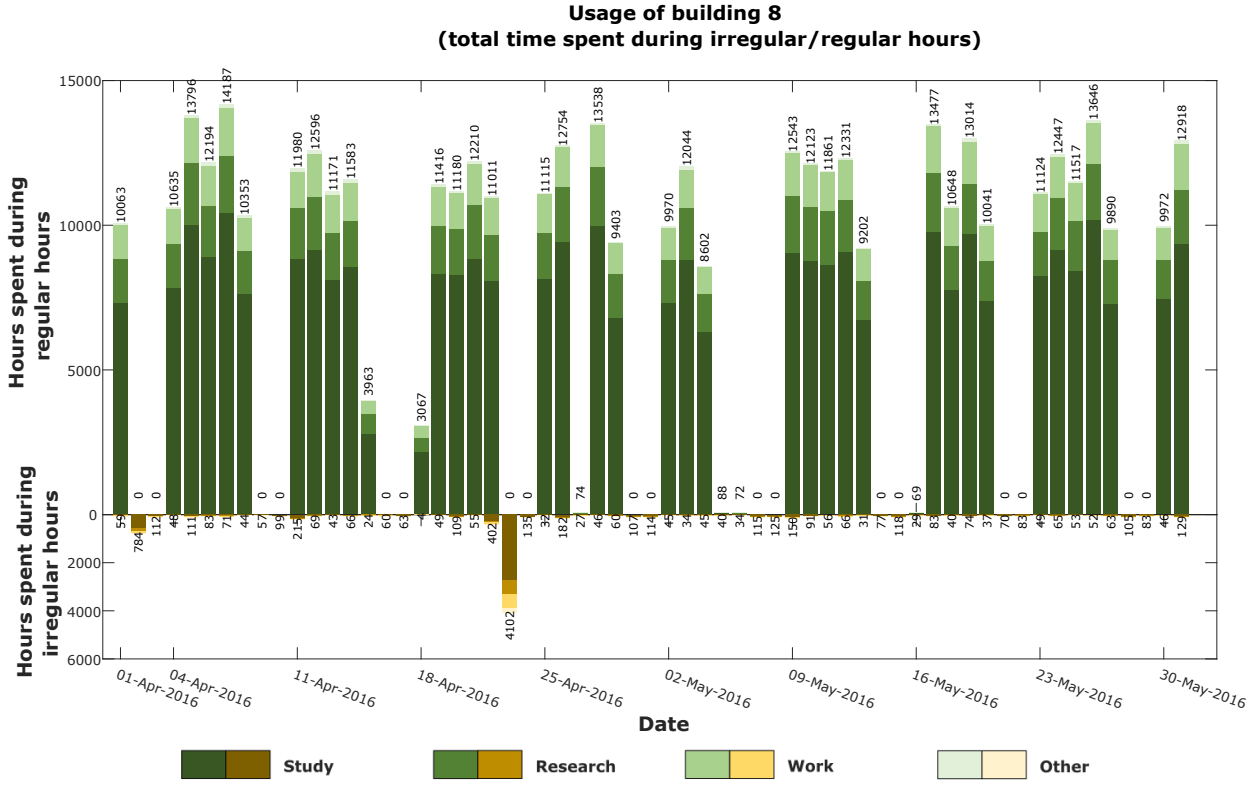


Figure A.11: Usage of building 8, the total time spent by users divided into activities

A.5. Building 12: Department of Chemical Engineering

Type of building: faculty building
Opening hours: Monday to Friday 07:30 - 18:00
Exceptions: authorised campuscard after closing

User number and time spent during irregular/regular hours
During the exam week, around 500 people use the building every weekday in regular hours (for 4.5 hours per person); around 90 people use it every weekend (for 4 hours per person).
After the exam, the building is more used in regular hours and the frequency of being used during irregular hours decreases. The daily user number in weekdays in regular hours varies from 475 to 868 but is about 500 (each person spends about 5.5 hours per day); the number for irregular hours varies from 107 to 219 but is around 120 on average (each person spends 1.5-2 hours per day). In general, there is a steady decrement of user number along with time after the start of the 4th quarter (all holidays are excluded). Main users of the building are students and the most common activity is study.

User group information
The building is mainly used by people from itself (in total 14,378 hours, on average 240 hours per days by all the users); therein, student (63%), academic staff (16.35%), support staff (14.75%) and others (5%). The building is barely used by people from other buildings; people from TU Delft library, Aula use the building for 7.5 hours, 3.5 hours per day on average respectively.

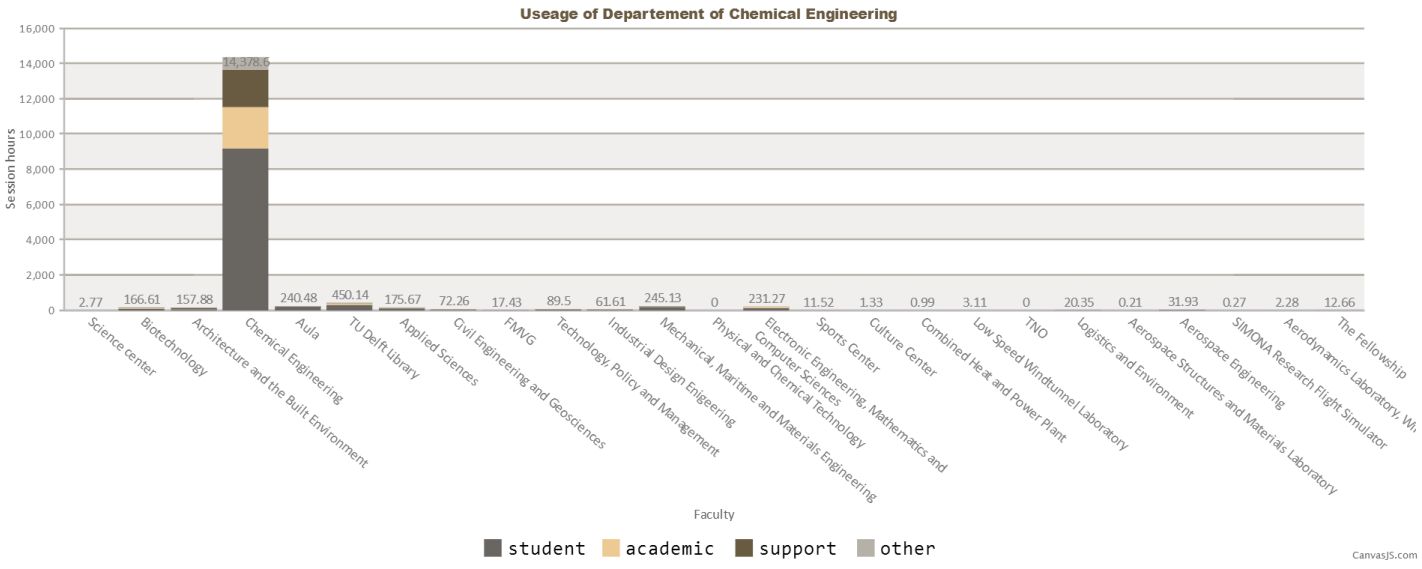


Figure A.12: Information of user group of building 12

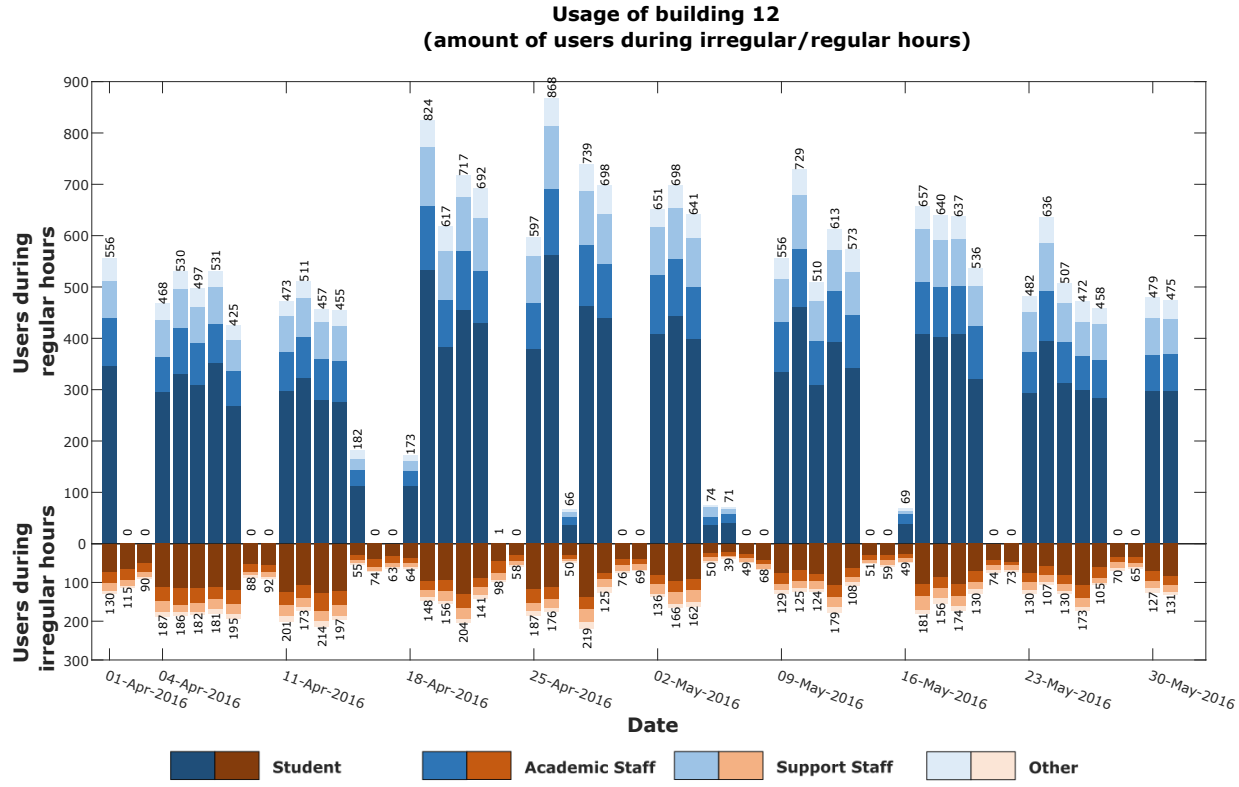


Figure A.13: Usage of building 12, amount of users present, divided by type of user

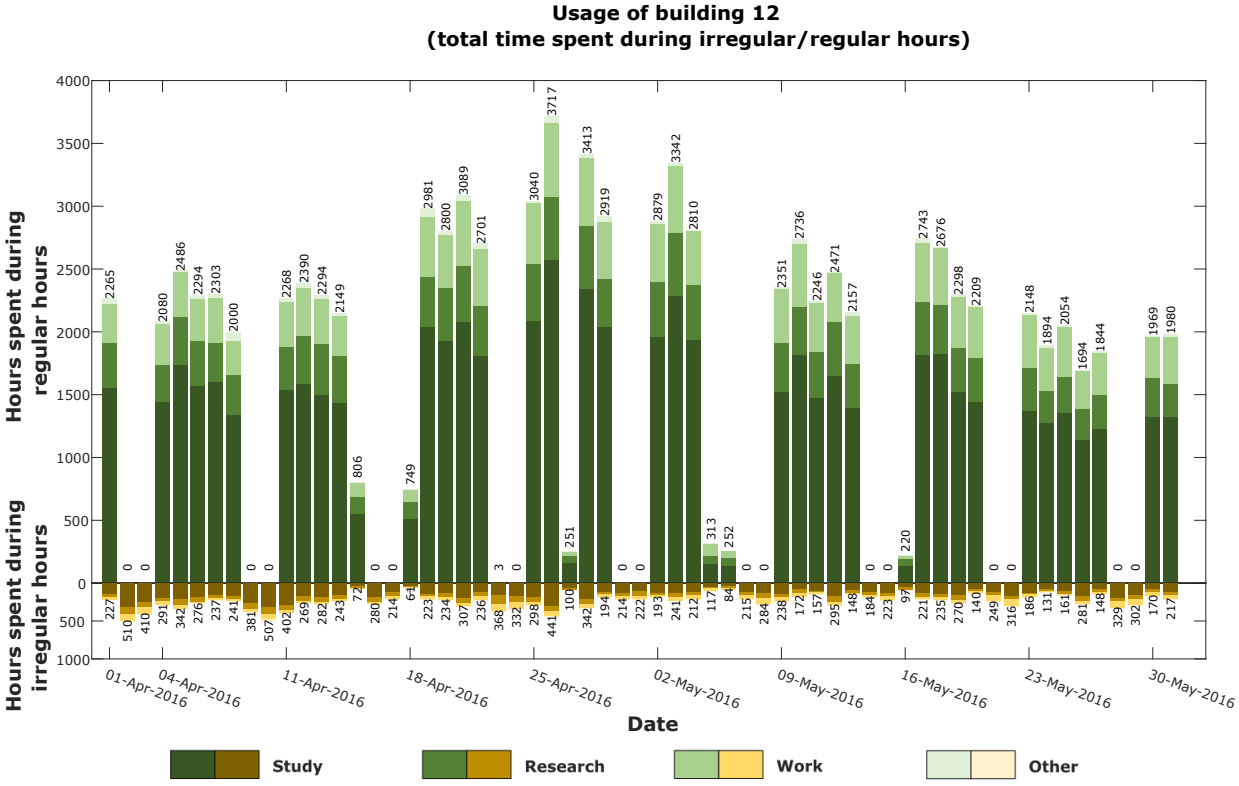


Figure A.14: Usage of building 12, the total time spent by users divided into activities

A.6. Building 20: Aula Conference Centre

Type of building: educational building
Opening hours: Monday to Friday 08:00 - 22:00
Exceptions: open during events (in weekend)

User number and time spent during irregular/regular hours
Figure A.16 shows that the amount of users during the irregular hours during the weekends is much larger than the amount of users during the irregular hours on weekdays. The number of users during the irregular hours on the 2nd, 3rd, 9th, and 10th of April are much higher than the number of users during the other Saturdays and Sundays of the two month period.
On Fridays there are always less people using the Aula than on the other normal weekdays. The Aula is used by many people on Tuesday 19th of April and Thursday 21st of April.

The large amount of user and hours spent on the 2nd, 3rd, 9th, and 10th of April are indicating that the Aula actually was not closed but opened. As this increased use in during the exam period it is assumed that the Aula is opened to provide students extra places to study for their exams.
The large amount of users and hours spent on Tuesday 19th of April can be explained because there was a minor market organised at the Aula which many people visited to gather information about available minors. On Thursday 21st of April the TU Delft organised a master event at the Aula, which was visited by many people as can be derived from the figures.

User group information
Aula is used by people from various buildings; people from the library use Aula for the longest time (in total 6,988 hours, on average 116.5 hours per days by all the users; therein, student (61%), academic staff (16.41%), support staff (14.4%) and others (8%)). Aula was used by people from itself for 4,135 hours in total (in which 75% by the students, 5.93% by academic staff, 7.18% by support staff and 11.9% by others). Aula is also used by people from main faculties; in which people from Applied sciences, 3ME, Architecture and Built Environment, Civil Engineering and Geosciences, Industrial Design Engineering and Technology, Policy and Management use the building for 33.7 hours, 22 hours, 16 hours, 15 hours, 15 hours and 14.4 hours per day by all the users on average. Aula is a comprehensive building which provides both lecture rooms and catering for people from all faculties and guests; perhaps this will be the reason explains why Aula is used by complex user groups.

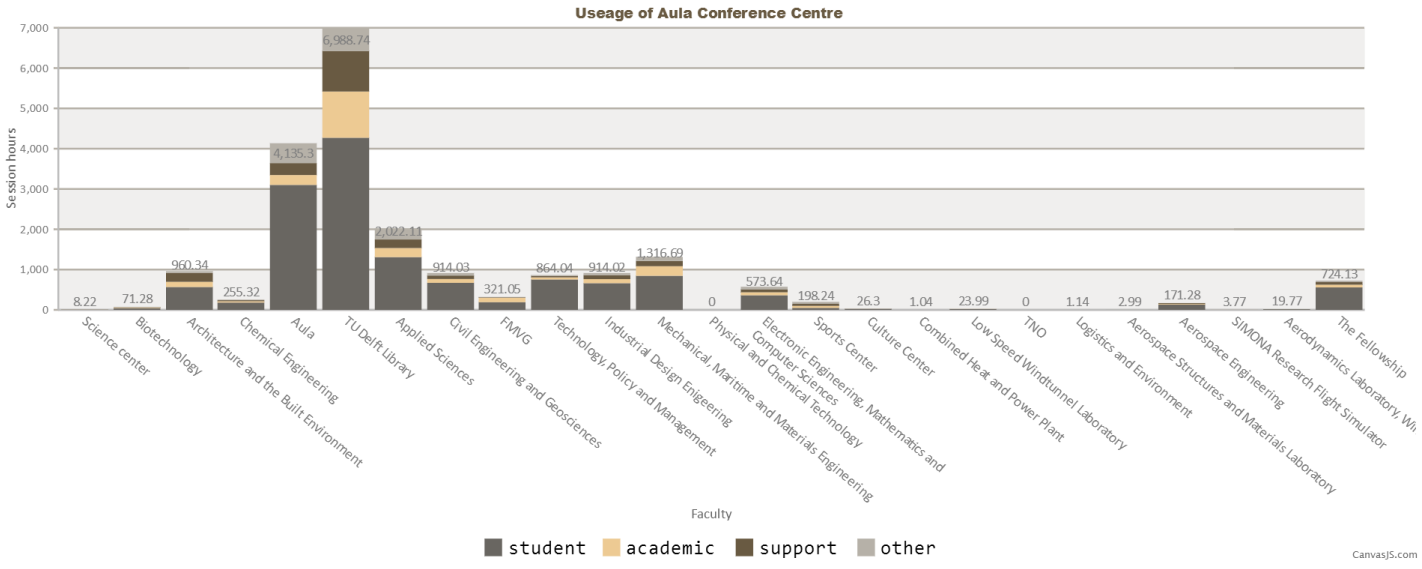


Figure A.15: Information of user group of building 20

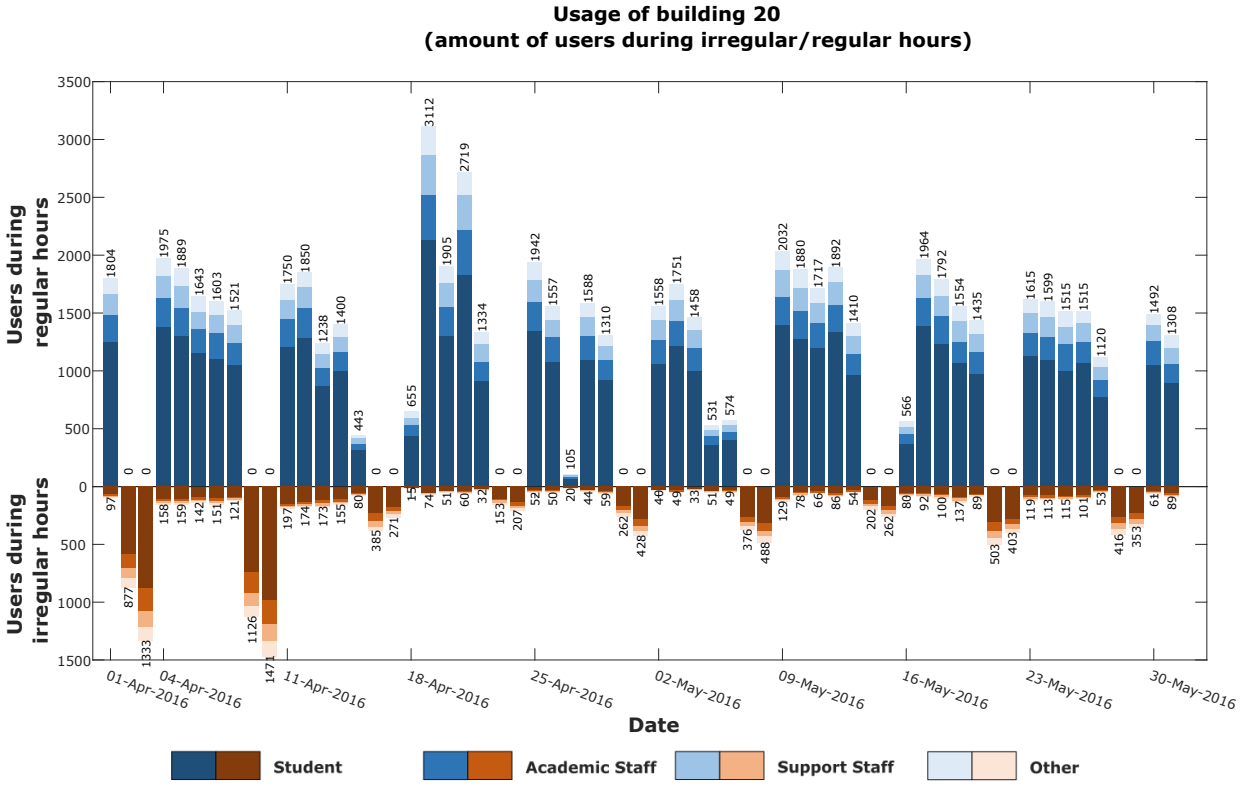


Figure A.16: Usage of building 20, amount of users present, divided by type of user

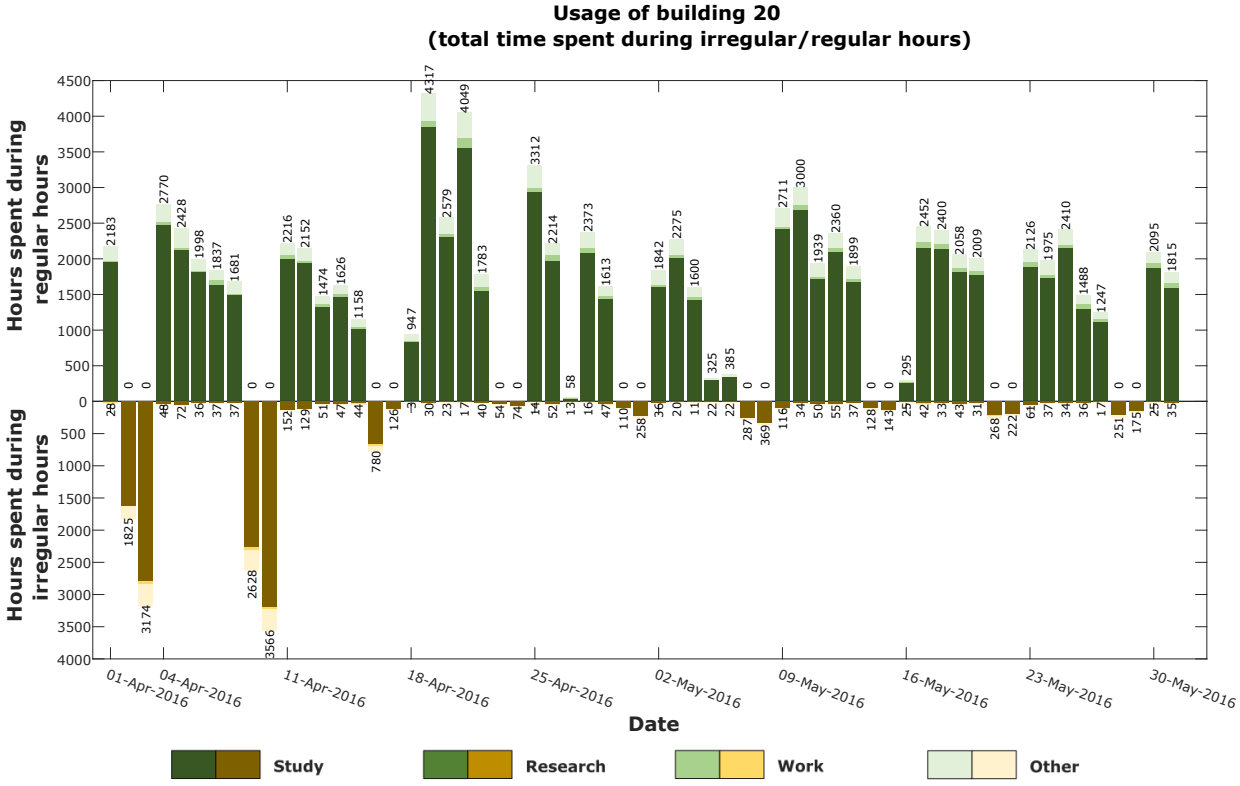


Figure A.17: Usage of building 20, the total time spent by users divided into activities

A.7. Building 21: TU Delft Library

Type of building: educational building
Opening hours: Monday to Sunday 08:00 - 24:00
Exceptions: exam period opening hours 08:00 - 02:0

User number and time spent during irregular/regular hours
Figure A.19 shows that the TU Delft Library is used the most from the 1st of April until the 14th of April during the regular and irregular hours. After this period there are always less than 30 people using the eduroam network of the TU Delft Library during the irregular hours. One thing that can be noticed is that the amount of users on Friday is always lower compared to the other weekdays. The TU Delft Library is used less in weekends compared to working days, and there are always more people using the TU Delft Library on Sundays than on Saturdays.
As can be seen in Figure A-13 the largest group of users of the TU Delft Library are students, followed by academic staff, support staff, and others.

Total time during irregular/regular hours
Figure A.20 shows that most people are using the TU Delft Library to study. Some time spend at the TU Delft Library is used for work or other activities.

User group information
TU Delft Library is used by people from various buildings; people from itself use it for the longest time (in total 38,751 hours, on average 646 hours per days by all the users; therein, student (68%), academic staff (15.3%), support staff (7.53%) and others (9%)). The library is also frequently used by people from main faculties; in which people from Civil Engineering and Geosciences, Architecture and Built Environment, Mechanical, Maritime and Material Engineering, Electronic Engineering, Mathematics and Computer use the building for 57 hours, 56 hours, 51 hours and 32.6 hours per day by all the users on average. Students spend a long period (longer than the time they spend in other buildings) doing project or thesis in the library will be classified into people from the library although they belong to their own faculty.

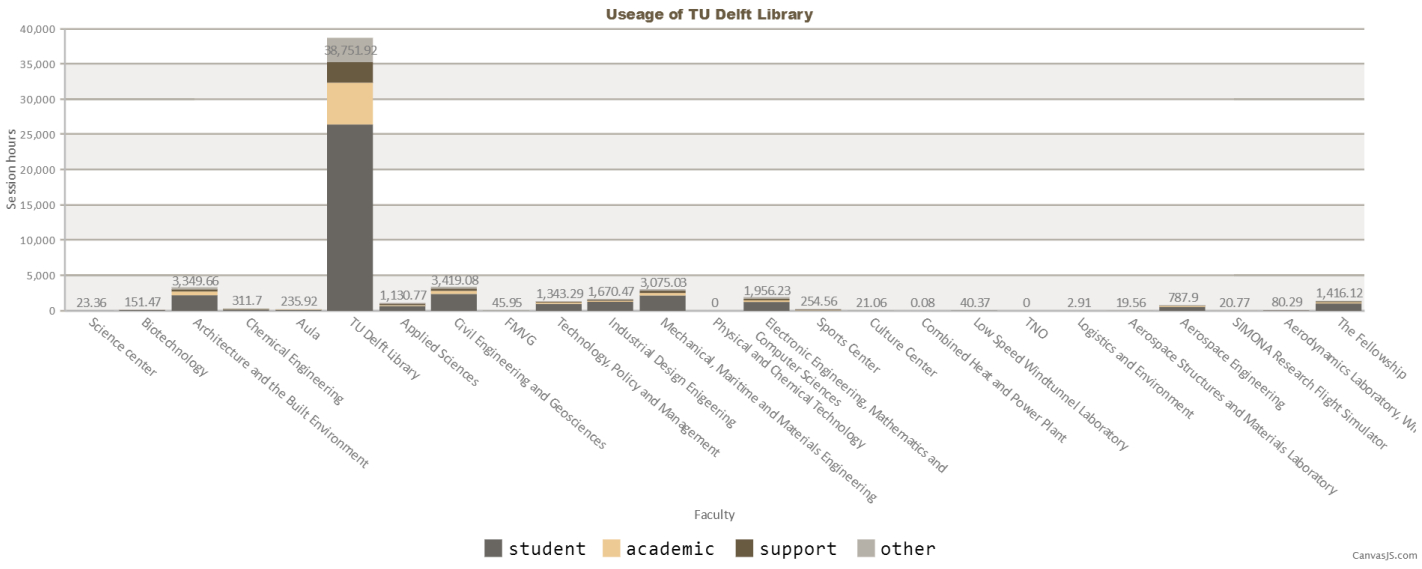


Figure A.18: Information of user group of building 21

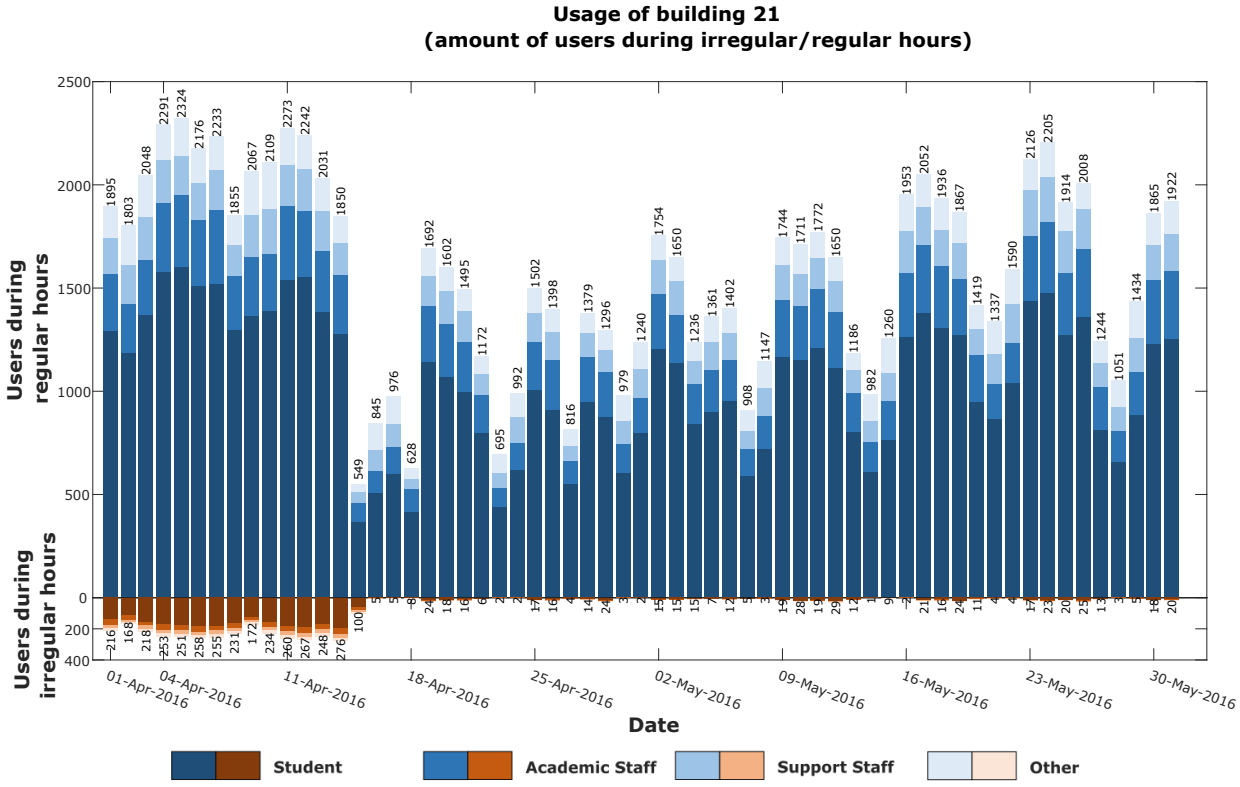


Figure A.19: Usage of building 21, amount of users present, divided by type of user

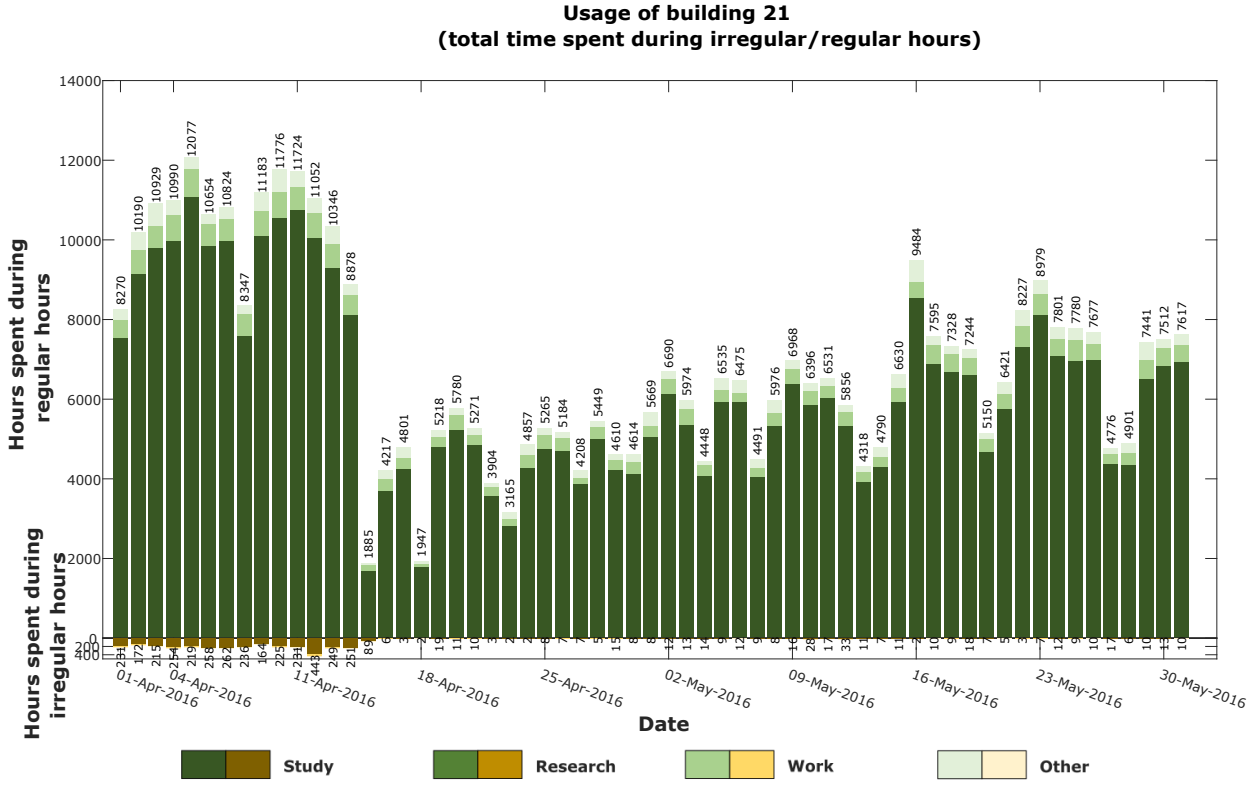


Figure A.20: Usage of building 21, the total time spent by users divided into activities

A.8. Building 22: Faculty of Applied Sciences

Type of building: faculty building
Opening hours: Monday to Friday 07:00 - 18:00
Exceptions: authorised campuscard after closing

User number and time spent during irregular/regular hours
Department of Applied Science is more use during irregular hours during exam week than after the exam. During the exam week, around 400 people use the building every weekday in irregular hours (for 1.5-2 hours per person); 200-300 people use it every weekend (for 3 hours per person).
After the exam, the building is more used in regular hours and the frequency of being used during irregular hours decreases. The daily user number in weekdays in regular hours varies from 804 to 1789 but is around 1400 on average (each person spends about 3.5 hours per day); the number for irregular hours varies from 179 to 488 but is 320 on average (each person spends 1.5 hours per day). Main users of the building are students and the most common activity is study. In general, there is a steady decrement of user number along with time after the start of the 4th quarter (all holidays are excluded). Main users of the building are students and the most common activity is study.

User group information
The building is mainly used by people from itself (in total 26,433 hours, on average 440 hours per days by all the users); therein, student (66%), academic staff (13.84%), support staff (13.69%) and others (6.5%). The building is barely used by people from other buildings; people from TU Delft library, Mechanical, Maritime and Material Engineering use the building for 24.5 hours, 12 hours per day on average respectively.

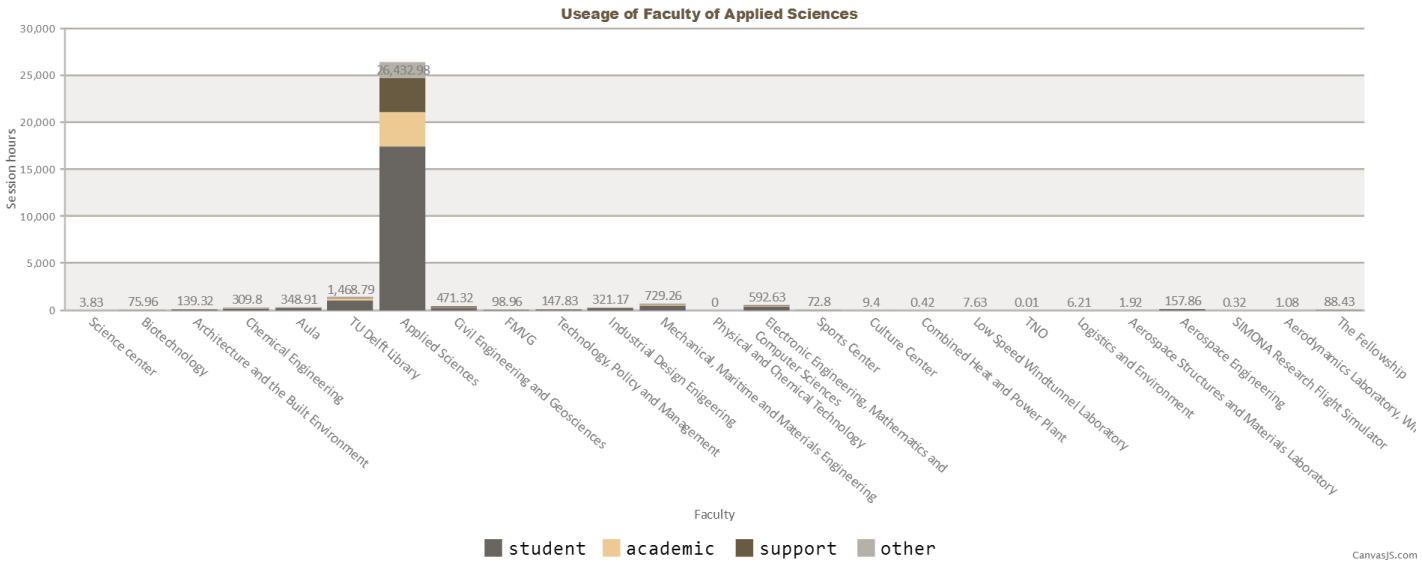


Figure A.21: Information of user group of building 22

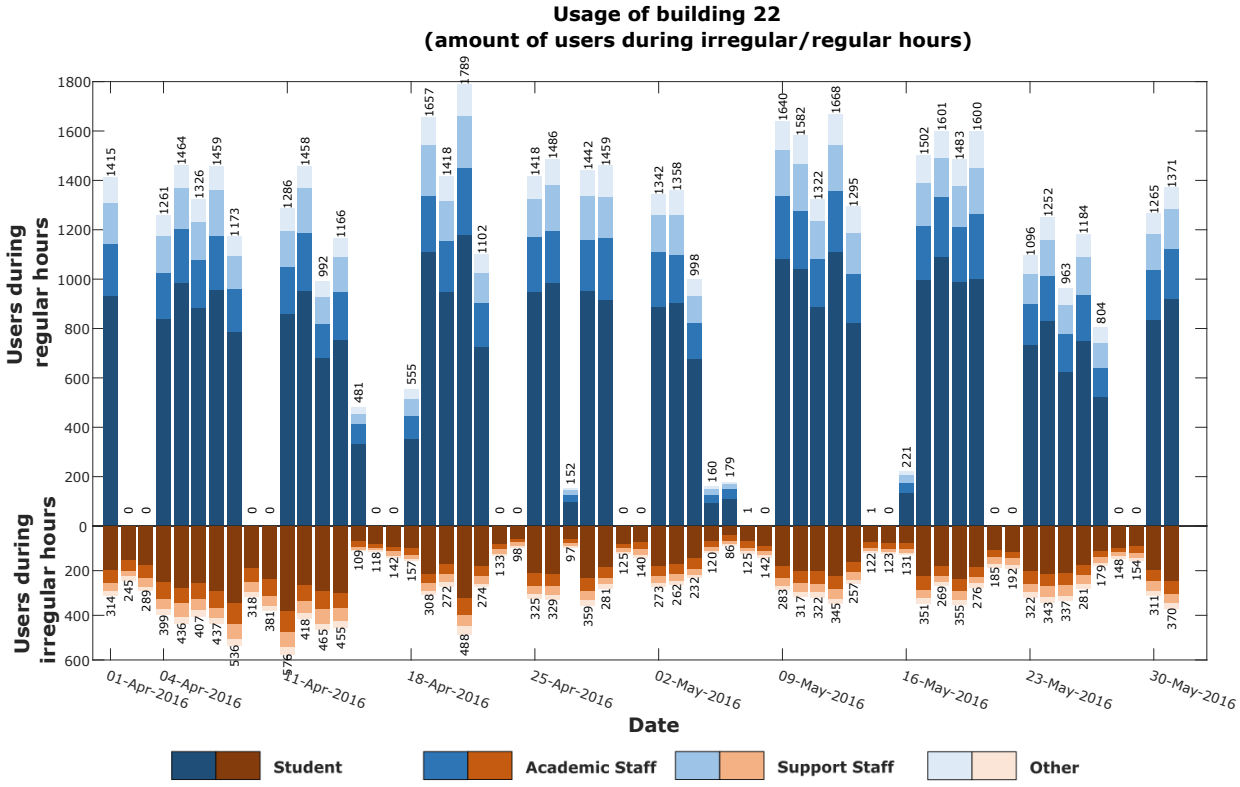


Figure A.22: Usage of building 22, amount of users present, divided by type of user

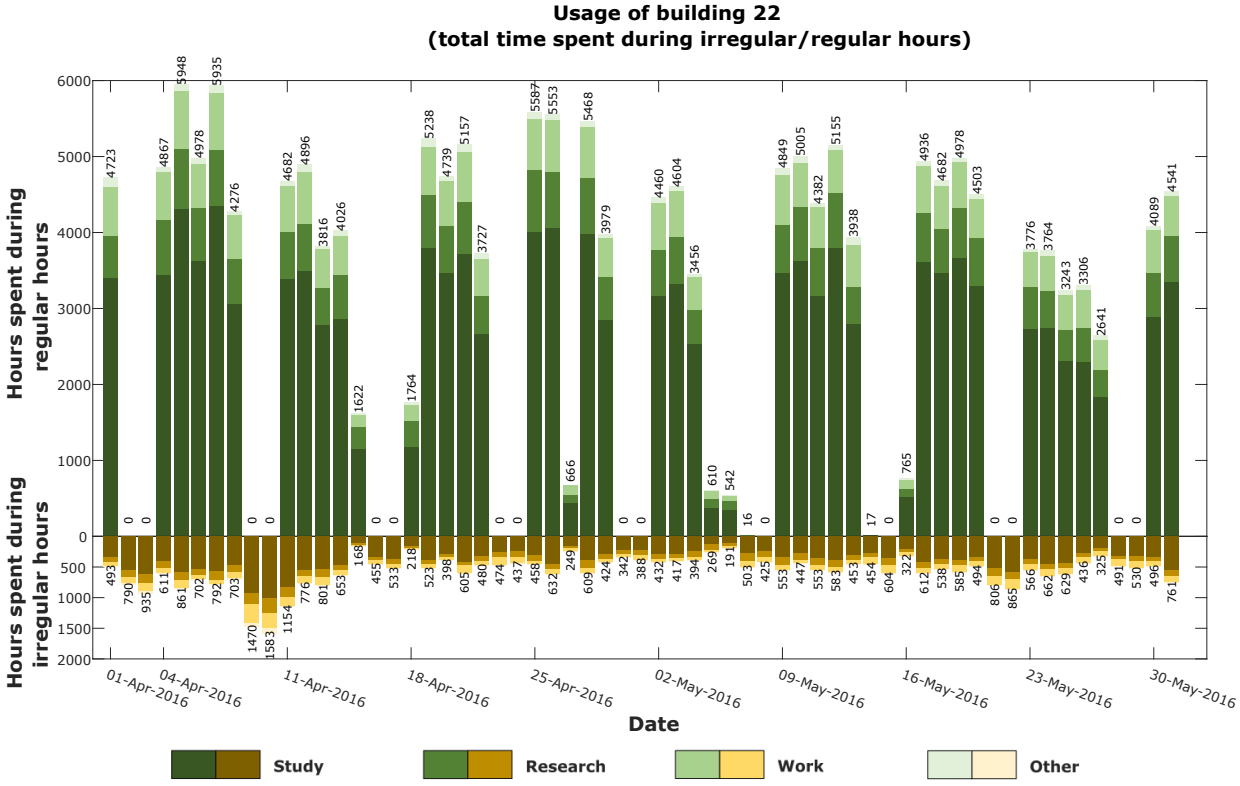


Figure A.23: Usage of building 22, the total time spent by users divided into activities

A.9. Building 23: Faculty of Civil Engineering and Geosciences

Type of building: faculty building
Opening hours: Monday to Friday 07:00 - 23:00 and Saturday 08:30 - 18:00
Exceptions: none

User number and time spent during irregular/regular hours
Faculty of Civil Engineering and Geosciences is more used in regular hours after exam week than during the exam. During the exam week, around 2200 people use the building every weekday in regular hours (for 4.5 hours per person); around 80 people use it every weekend (for 3-4 hours per person).
After the exam, the daily user number in weekdays in regular hours varies from 2211 to 3404 but is around 2900 on average (each person spends about 4 hours per day); there is an obvious decrement of user number on every Friday. The user number for irregular hours varies from 21 to 193 but is 60-70 on average (each person spends 1.5 hours per day). Main users of the building are students and the most common activity is study. Compared with the other educational buildings, the building is less used in irregular hours.

User group information
The building is mainly used by people from itself (in total 52,293 hours, on average 870 hours per days by all the users); therein, student (62.6%), academic staff (16.3%), support staff (13.6%) and others (7.5%). The building is barely used by people from other buildings; people from TU Delft library, Mechanical, Maritime and Material Engineering use the building for 45 hours, 21 hours per day on average respectively.

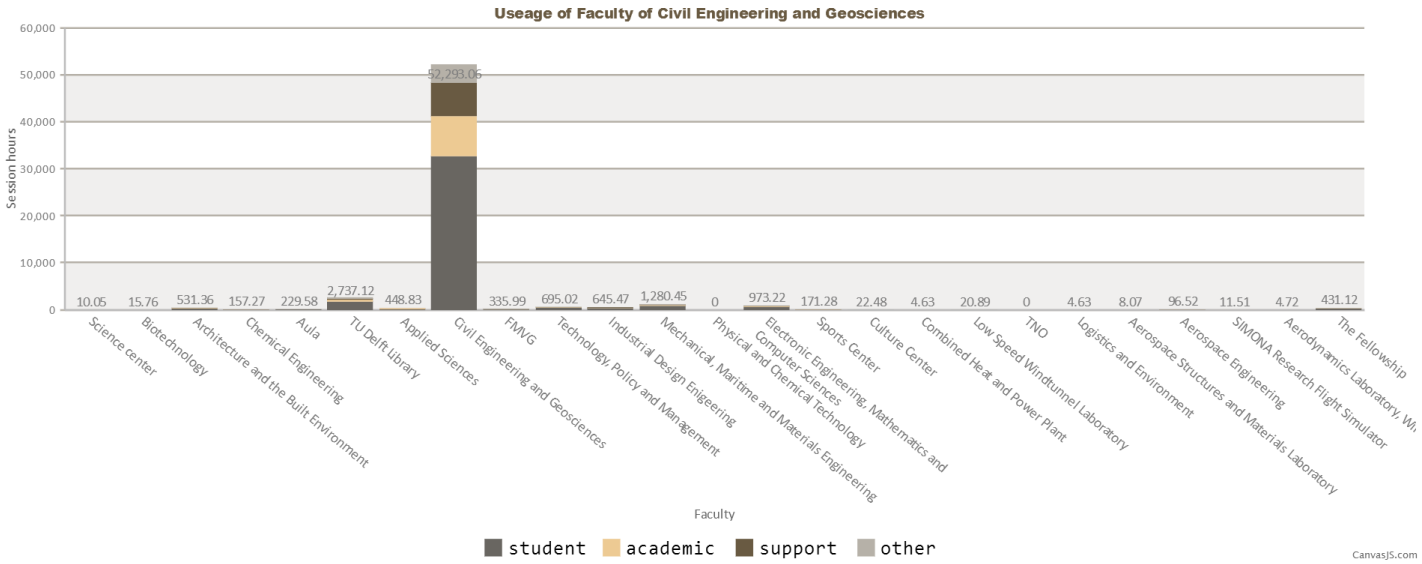


Figure A.24: Information of user group of building 23

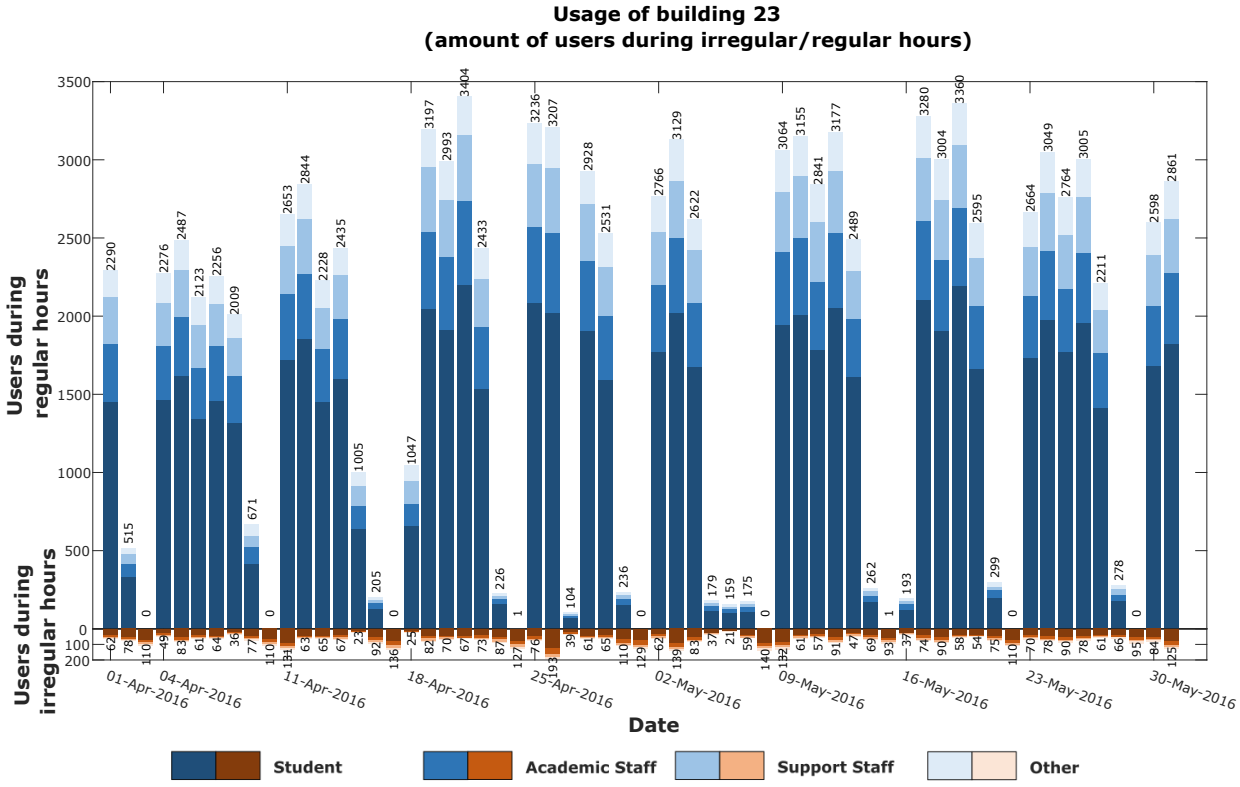


Figure A.25: Usage of building 23, amount of users present, divided by type of user

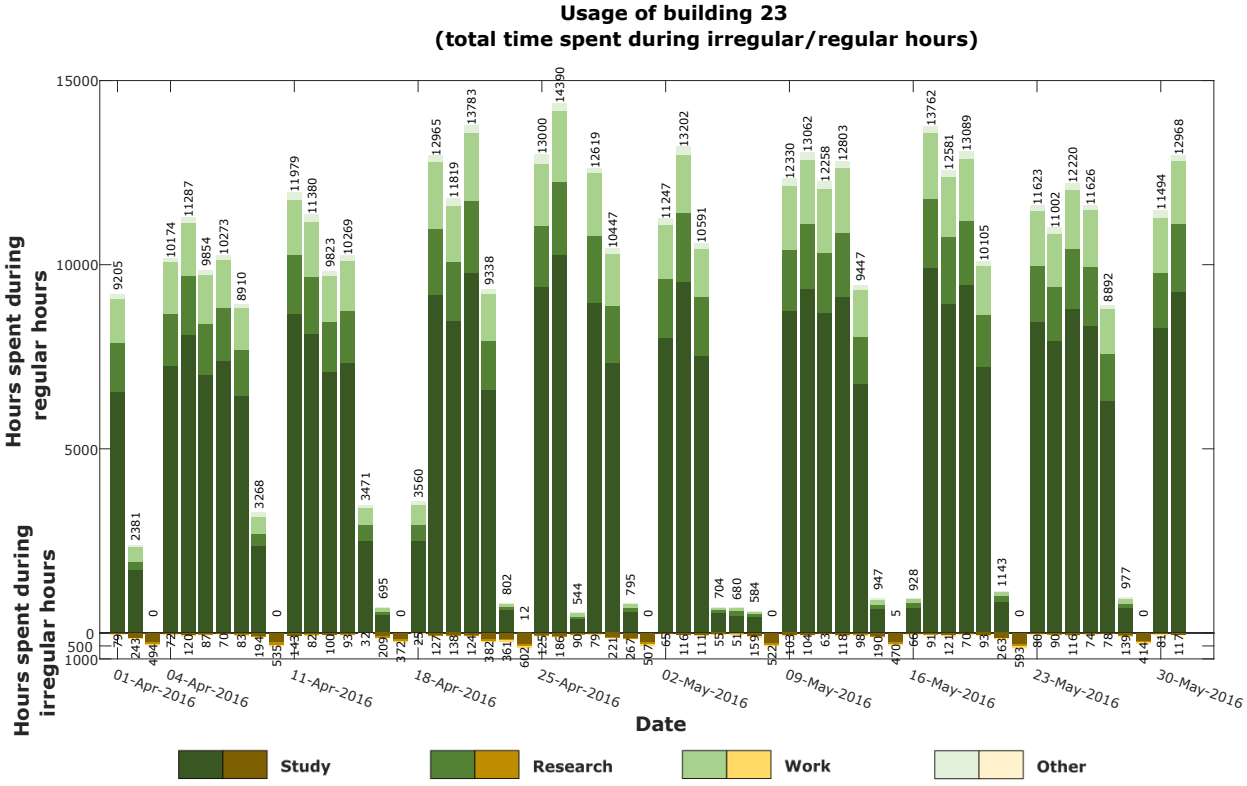


Figure A.26: Usage of building 23, the total time spent by users divided into activities

A.10. Building 30: Education and Student Affairs / FMRE

Type of building: facility management and real estate building
Opening hours: Monday to Friday 07:30 - 17:00
Exceptions: none

User number and time spent during irregular/regular hours
There is no obvious change of user number during the exam or after the exam in either regular hour or irregular hours. However, the maximum occupancy of the building also occurred during the first week in the 4th quarter like what happened for most of the campus buildings. The daily user number in weekdays in regular hours varies from 251 to 472 but is around 350-400 on average (each person spends about 4 hours per day); there is an obvious decrement of user number on every Friday.
The user number in weekdays in irregular hours varies from 104 to 201 and there is also an obvious decrement on every Friday (each person spends less than 1 hour per day). Main users of the building are students and the most common activity is undefined errands and working due to the truth that many students visit here and many support staff work in here for a longer time.

User group information
The building is mainly used by people from itself (in total 10,595 hours, on average 176 hours per days by all the users); therein, student (69%), academic staff (10%), support staff (10.5%) and others (10.5%). The building is barely used by people from other buildings; people from Applied Sciences for 2.5 hours per day on average respectively.

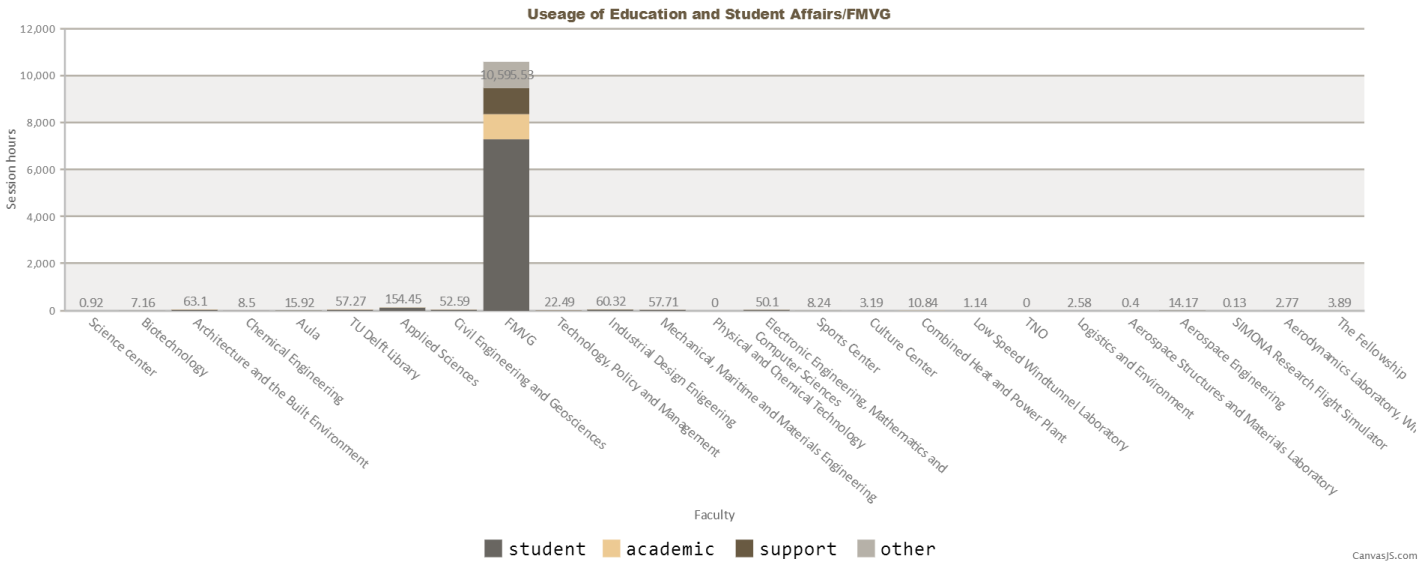


Figure A.27: Information of user group of building 30

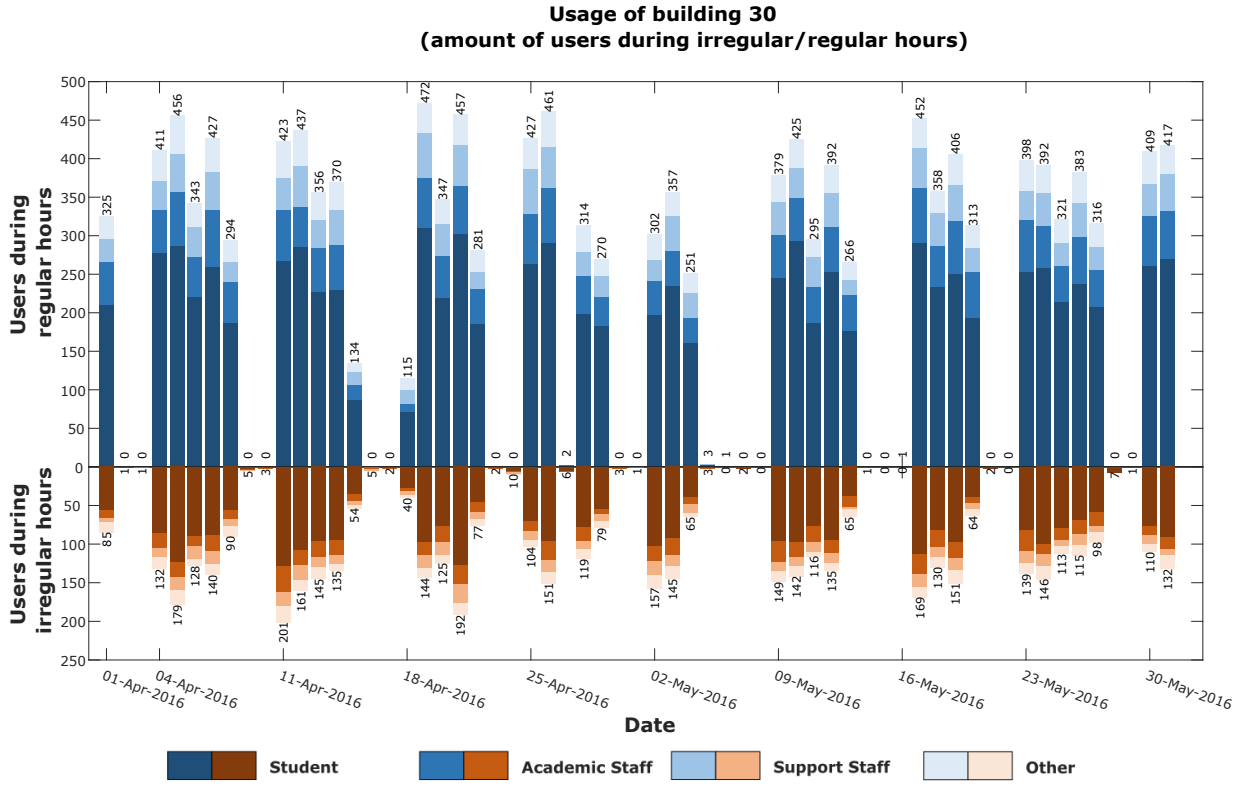


Figure A.28: Usage of building 30, amount of users present, divided by type of user

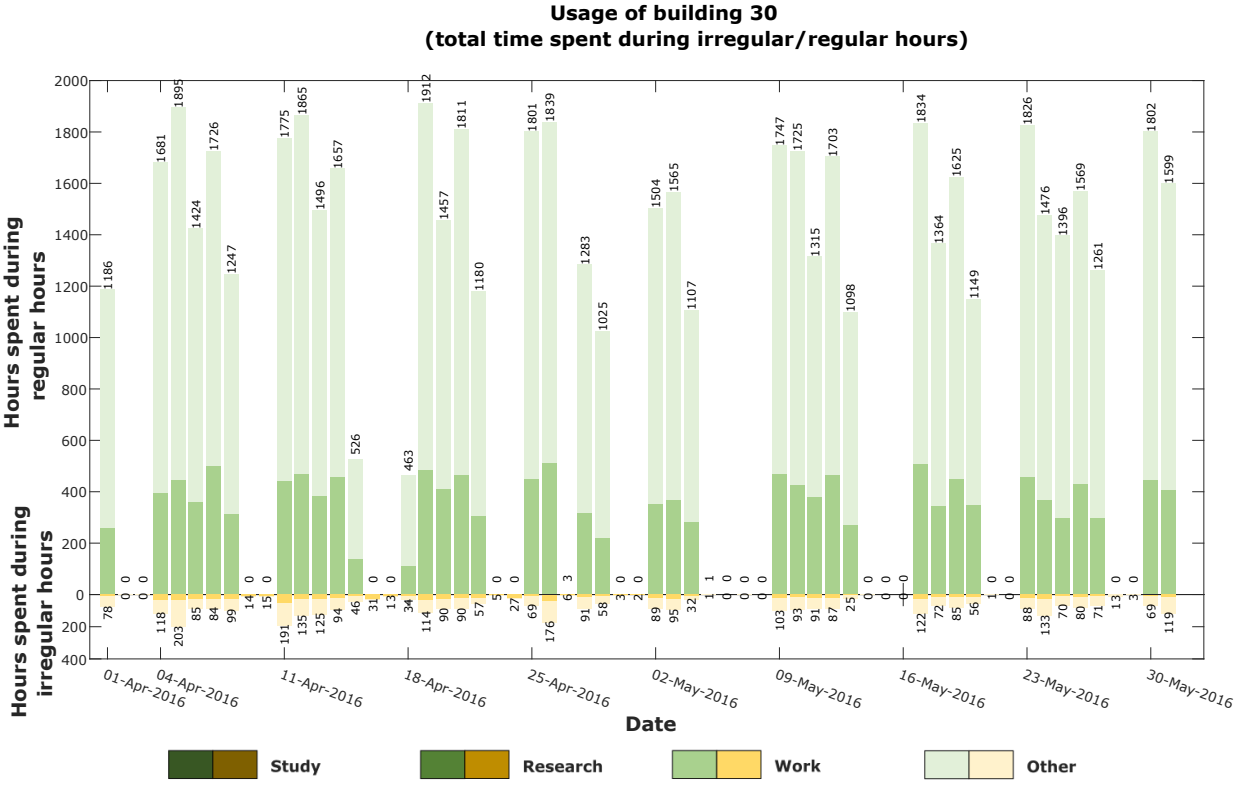


Figure A.29: Usage of building 30, the total time spent by users divided into activities

A.11. Building 31: Faculty of Technology, Policy and Management

Type of building: faculty building
Opening hours: Monday to Friday 07:30 - 22:30
Exceptions: weekends 10:00 - 18:00 with authorised campuscard

User number and time spent during irregular/regular hours
Faculty of Technology, Policy and Management is more use during irregular hours during exam week than after the exam. During the exam week, the user number in weekdays decreases along with the time. The average user number is around 900 on every weekday in regular hours (for 4 hours per person); 40 people use it every weekend (for 2.5 hours per person).
After the exam, the building is more used in regular hours and the frequency of being used during irregular hours decreases. The daily user number in weekdays in regular hours is around 1000 on average (each person spends about 3.5 hours per day); the number for irregular hours is 15 on average (each person spends 1 hour per day) in weekdays and 20 in weekends (each person spends 3 hours per day). An obvious decrease of user number can be observed on every Fridays. Main users of the building are students and the most common activity is study.

User group information
The building is mainly used by people from itself (in total 13,731 hours, on average 228 hours per days by all the users); therein, student (63.75%), academic staff (16.3%), support staff (12.3%) and others (7.7%). The building is also used by people from various building for a relative short period; people from TU Delft library, Architecture and Built Environment, Civil Engineering and Geosciences, Industrial Design Engineering and Mechanical, Maritime and Material Engineering use the building for 21 hours, 13 hours, 10 hours, 9.5 hours and 9.5 hours per day on average respectively. The reason why the Faculty of Technology, Policy and Management is used by people from different buildings is that TBM is a new building with sufficient rooms which is sometimes used by other faculties as lecture rooms or exam rooms.

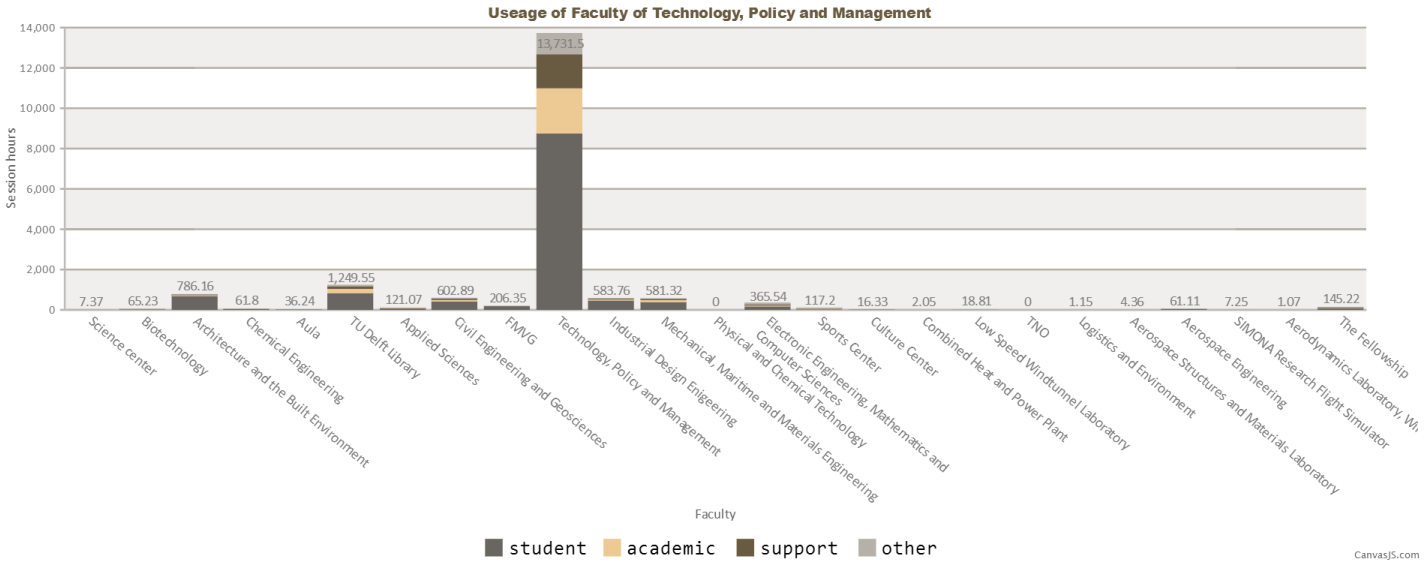


Figure A.30: Information of user group of building 31

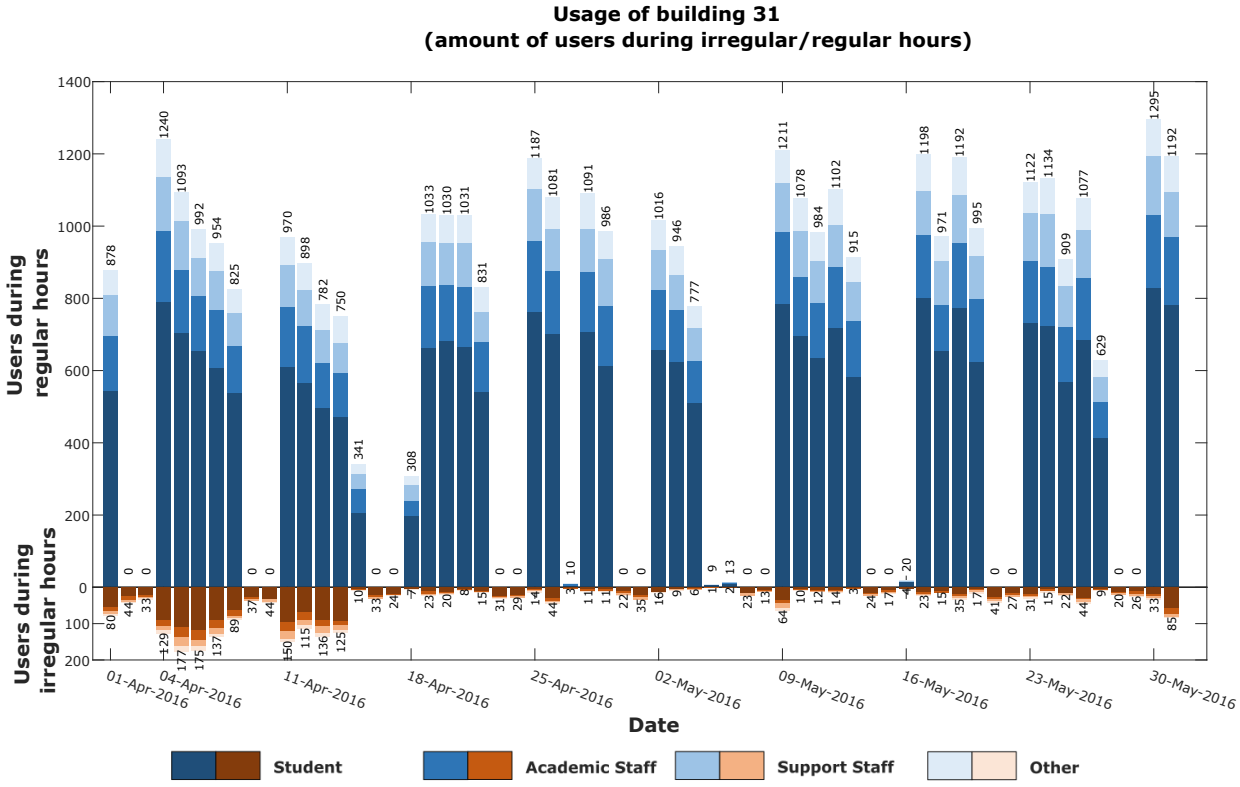


Figure A.31: Usage of building 31, amount of users present, divided by type of user

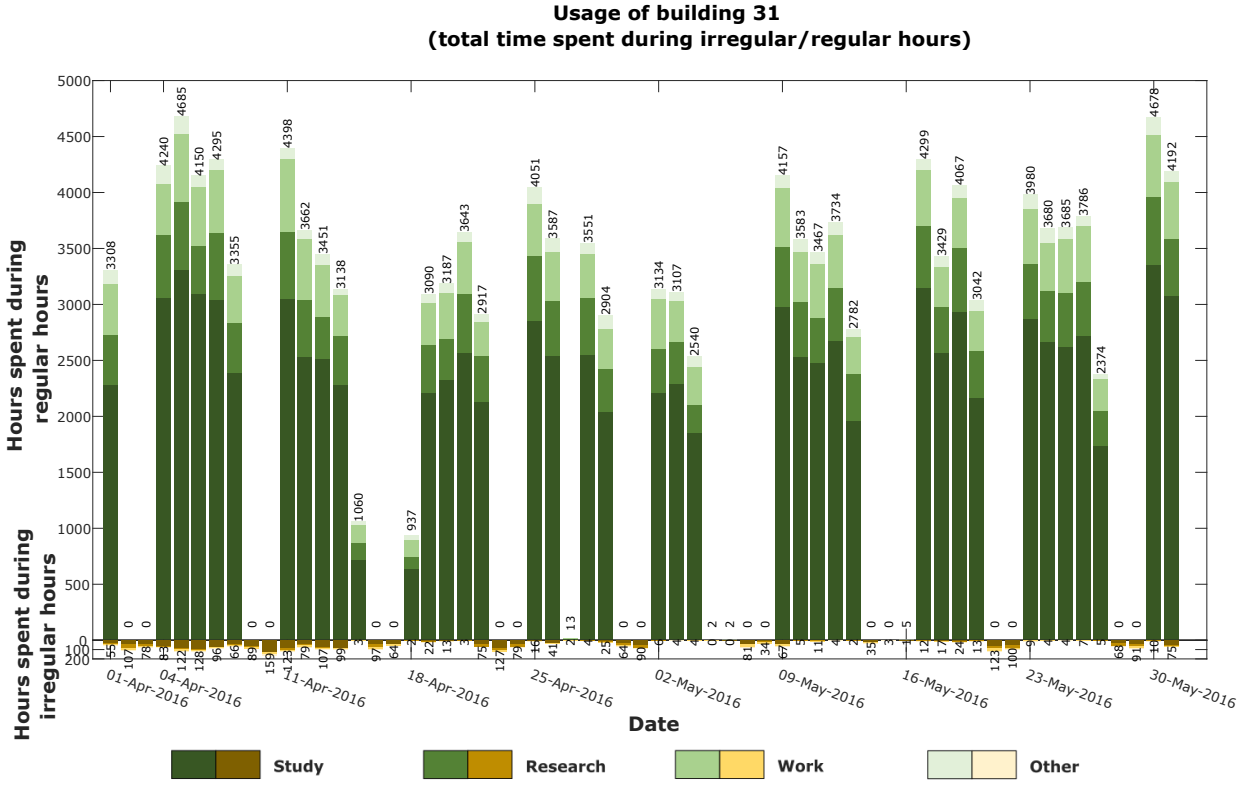


Figure A.32: Usage of building 31, the total time spent by users divided into activities

A.12. Building 32: Faculty of Industrial Design Engineering

Type of building: faculty building
Opening hours: Monday to Friday 08:00 - 22:00
Exceptions: 07:00 - 08:00 with authorised campuscard, Friday 08:00 - 19:00

User number and time spent during irregular/regular hours
Faculty of Industrial Design Engineering is more used in regular hours after exam week than during the exam. During the first exam week, around 2200 people use the building every weekday in regular hours (for 4.5 hours per person); during the second exam week, around 1600 people use the building every weekday in regular hours (for 4.5 hours per person). Around 40 people use it every weekend (for 4 hours per person).
After the exam, the daily user number in weekdays in regular hours is around 2600-2800 on average (each person spends about 3.5-4 hours per day); there is an obvious decrement of user number on every Friday. The user number in irregular hours varies from 30-290 but is 50-70 on average (each person spends 1-1.5 hours per day). Main users of the building are students and the most common activity is study. Compared with the other educational buildings, the building is less used in irregular hours.

User group information
The building is mainly used by people from itself (in total 44,363 hours, on average 740 hours per days by all the users); therein, student (68%), academic staff (13%), support staff (10.9%) and others (7.7%). The building is barely used by people from other buildings; however, it is used by people from Mechanical, Maritime and Material Engineering for a relatively longer period (9.5 hours per day on average by all users). The possible reason is that these two building are adjacent; therefore, some users are accidentally connected with access points in another building.

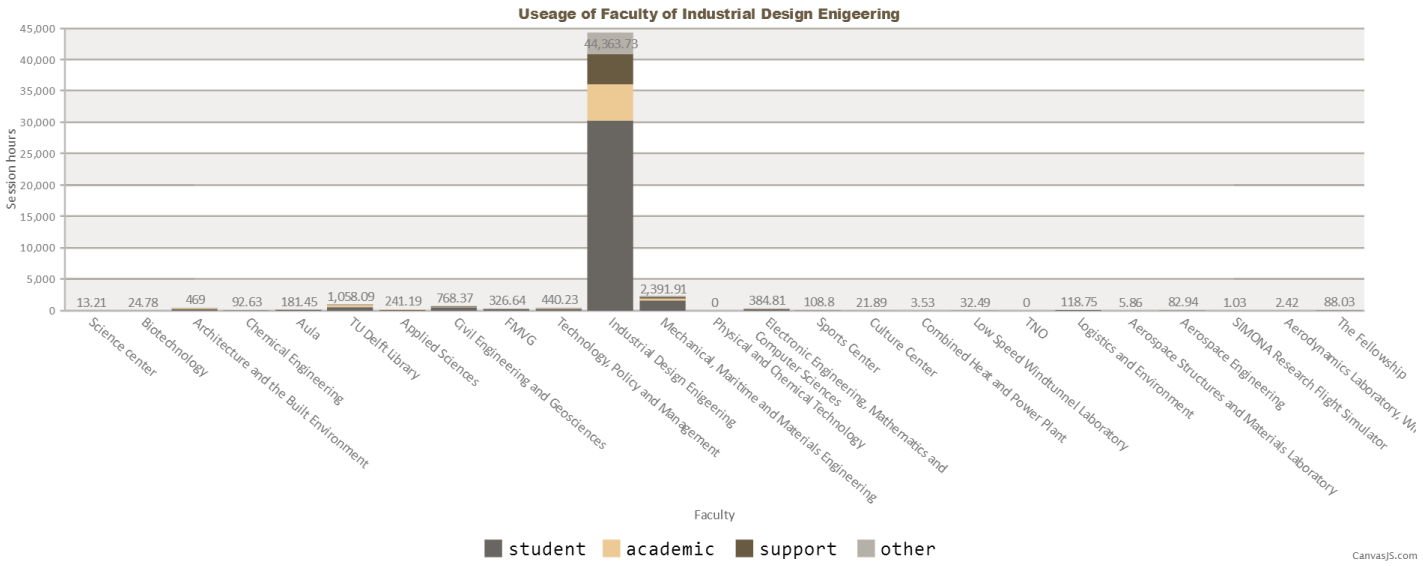


Figure A.33: Information of user group of building 32

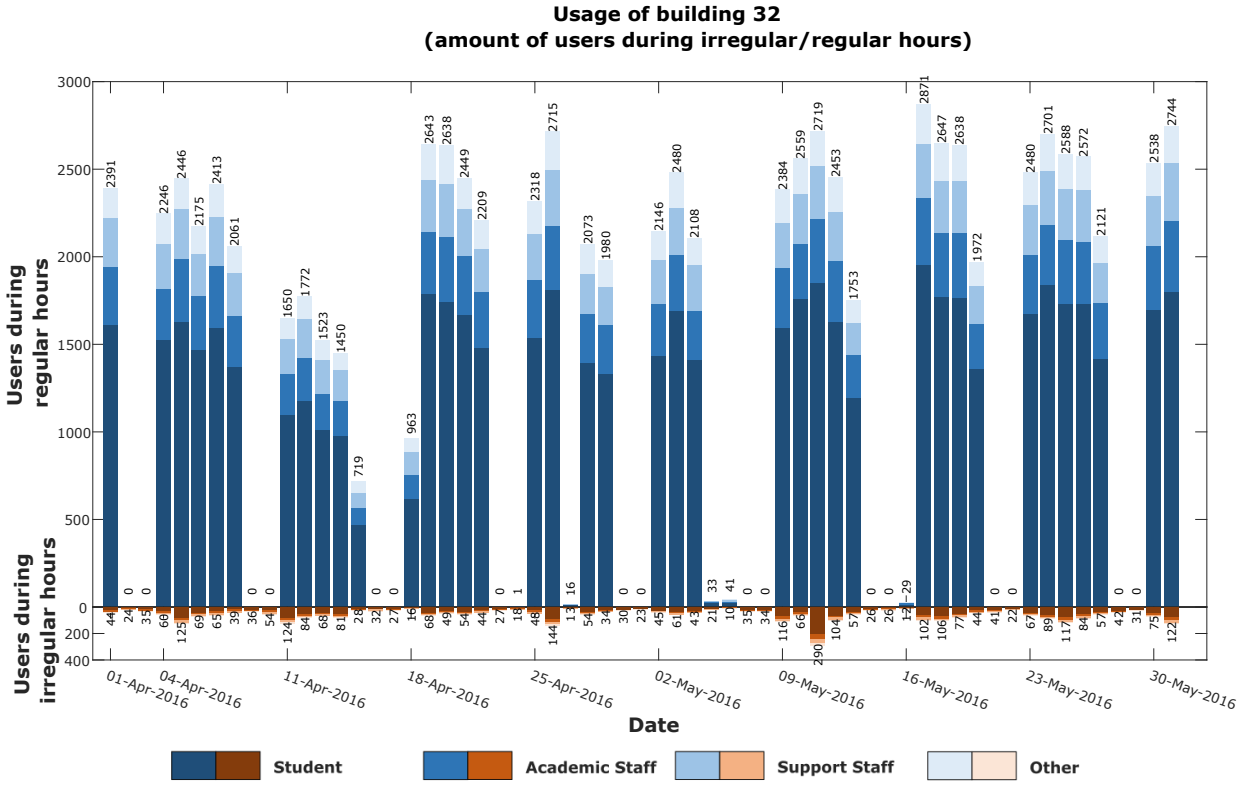


Figure A.34: Usage of building 32, amount of users present, divided by type of user

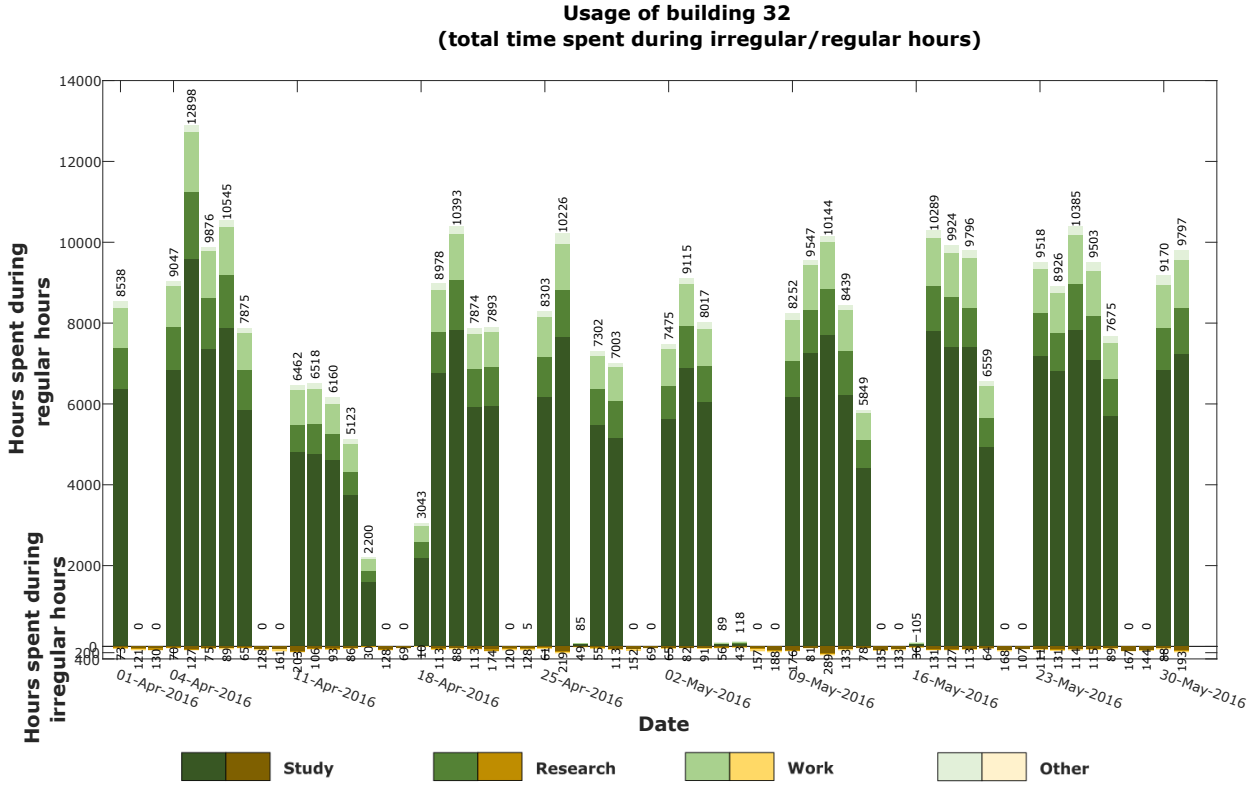


Figure A.35: Usage of building 32, the total time spent by users divided into activities

A.13. Building 34: Faculty Mechanical, Maritime and Materials Engineering

Type of building: faculty building
Opening hours: Monday to Friday 08:00 - 19:00
Exceptions: authorised campuscard after closing

User number and time spent during irregular/regular hours
The building is more used in regular hours after the exam and more used in irregular hours during exam week. The user number in regular hours is around 2000 during exam week; the number increases to around 2900 after the exam. The maximum occupancy of the building occurred on 9th May 2016, 21th April 2016 and 19th April 2016; latter two are in the first week of the 4th quarter, during which most the new courses started. Each person spends about 3 to 3.5 hours in regular hours per weekday.
In the irregular hours in weekdays during the exam week, the daily user number is around 600; the number decreases to around 300 after the exam. The time each person spends also decreases from 2 hours to 1.5 hours.
In the irregular hours in weekends during the exam week, the daily user number is around 270; the number decreases to around 90 after the exam. The time each person spends also decreases from 4-4.5 hours to 3-3.5 hours. Main users of the building are students and the most common activity is study.

User group information
The building is mainly used by people from itself (in total 45,501 hours, on average 758 hours per days by all the users); therein, student (63%), academic staff (15%), support staff (12%) and others (9.7%). The building is barely used by people from other buildings; however, it is used by people from Industrial Design Engineering for a relatively longer period (37.2 hours per day on average by all users). The possible reason is that these two building are adjacent; therefore, some users are accidentally connected with access points in another building. What is more, like what happens in other main faculties, the building is also frequently used by people from the library (61 hours per day on average by all users).

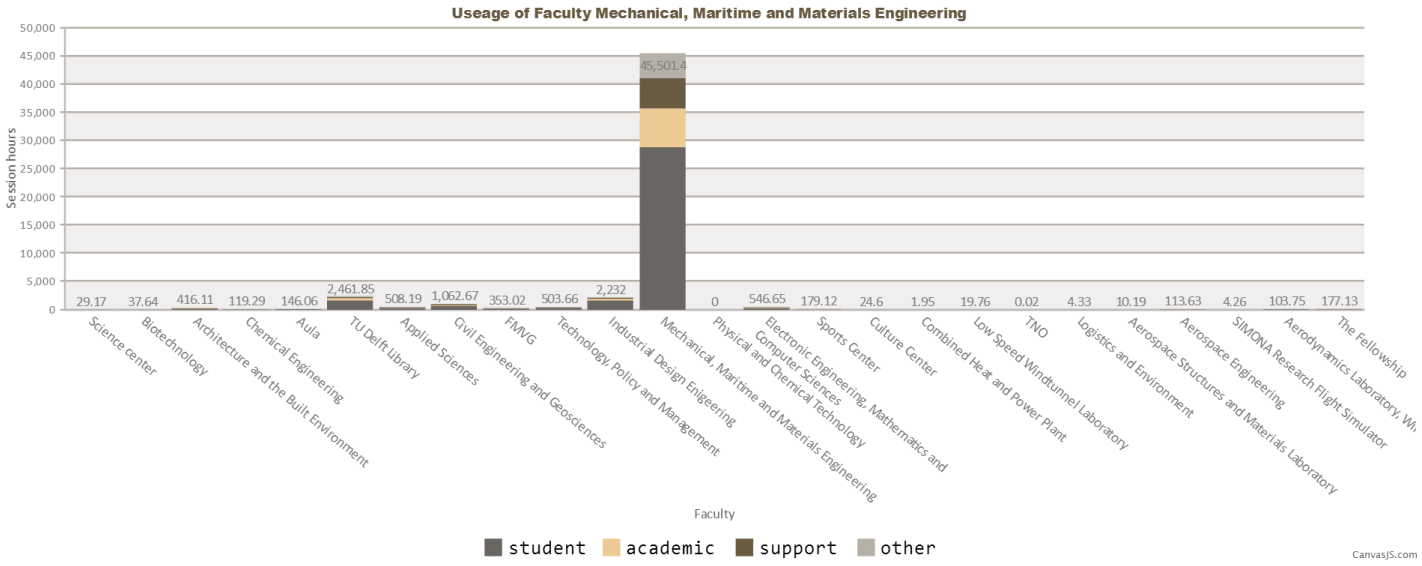


Figure A.36: Information of user group of building 34

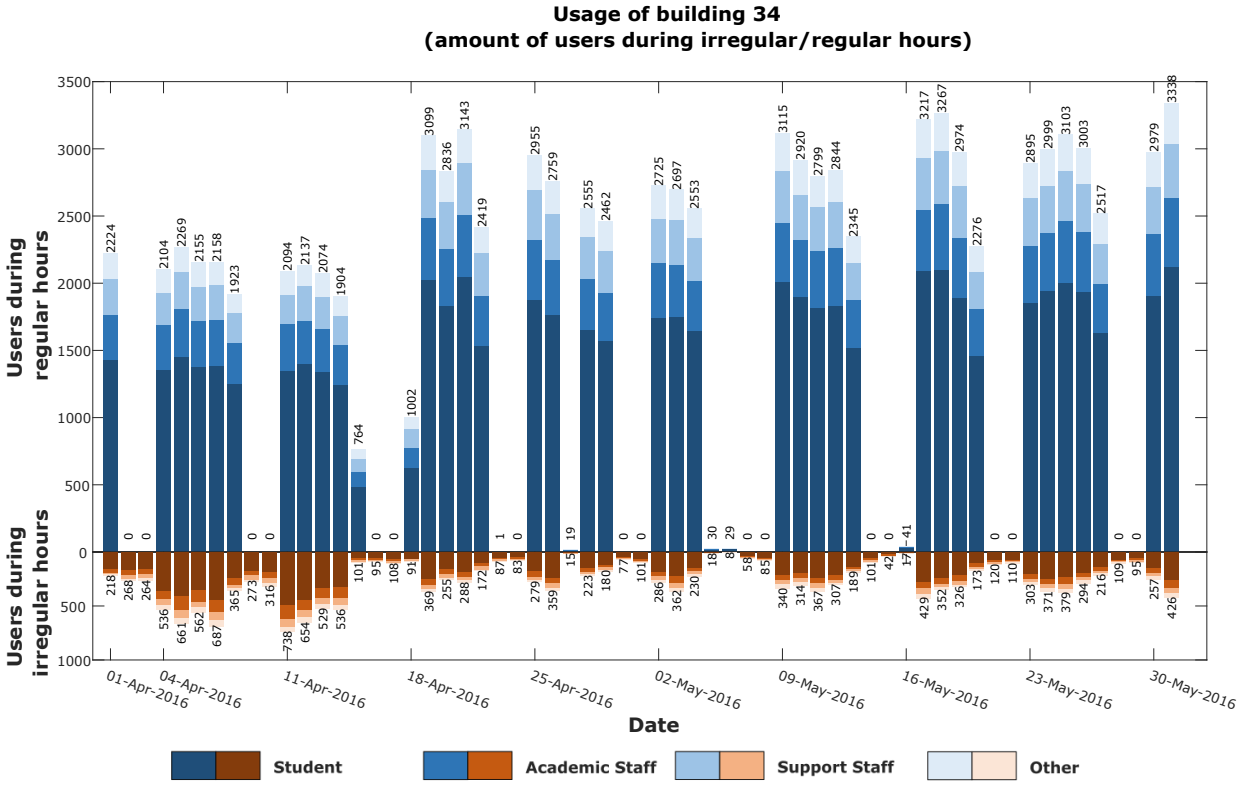


Figure A.37: Usage of building 34, amount of users present, divided by type of user

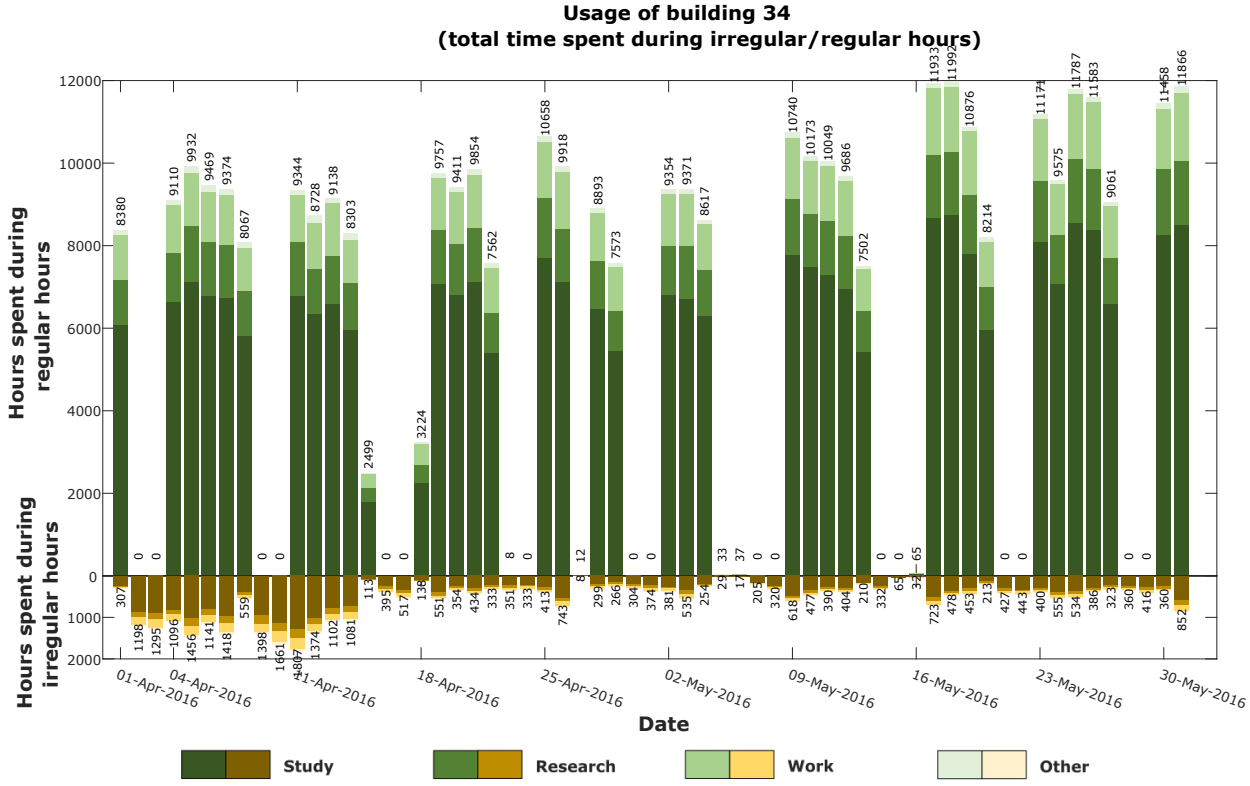


Figure A.38: Usage of building 34, the total time spent by users divided into activities

A.14. Building 35: Education Building 35

Type of building: educational building
Opening hours: Monday to Friday 07:30 - 18:00
Exceptions: none

User number and time spent during irregular/regular hours
The Educational building 35 is equally used after the exam week and during the exam. During weekdays in exam week, the daily user number in regular hours varies from 389 to 747 but the number is around 500 in general; each person spends about 1.5 hours per weekday in regular hours. The user number on the last day of exam week and the first day of the 4th quarter is 243 and 63 respectively.
After the exam, the daily user number of the building during regular hours in weekdays slight decreased to around 450 (on average). The maximum occupancy of the building occurred on 24th May 2016 which was 883 people. Although the average daily user number decreases, the average time people spend in the building increases to 2.5 hours.
During the irregular hours in weekdays, the building is more used during the exam week and after 9th May 2016 (on average 150 to 200 users per weekday) than the duration in between (varies from 10 to 50 users per day, on average 25 users per weekday); each user spends from 1.5 hours to 3 hours per day. During the regular hours in weekend, user number decreases to around 2 and the time each spends in the building also decreases to less than 1 hour.
Main users of the building are students and the most common activity is study.

User group information
Except people from Education Building 35 itself, it is frequently used by people from Electronic Engineering, Mathematics and Computer Science (32 hours per day on average by all users). The building is also frequently be used by people from main faculties; therein people from Mechanical, Maritime and Material Engineering, Civil Engineering and Geosciences and Architecture and Built Environment use the building for 80 hours, 4.4 hours and 3.3 hours per day by all the users on average. People who spend long time in the library also go to the building (10 hours per day by all the users on average).

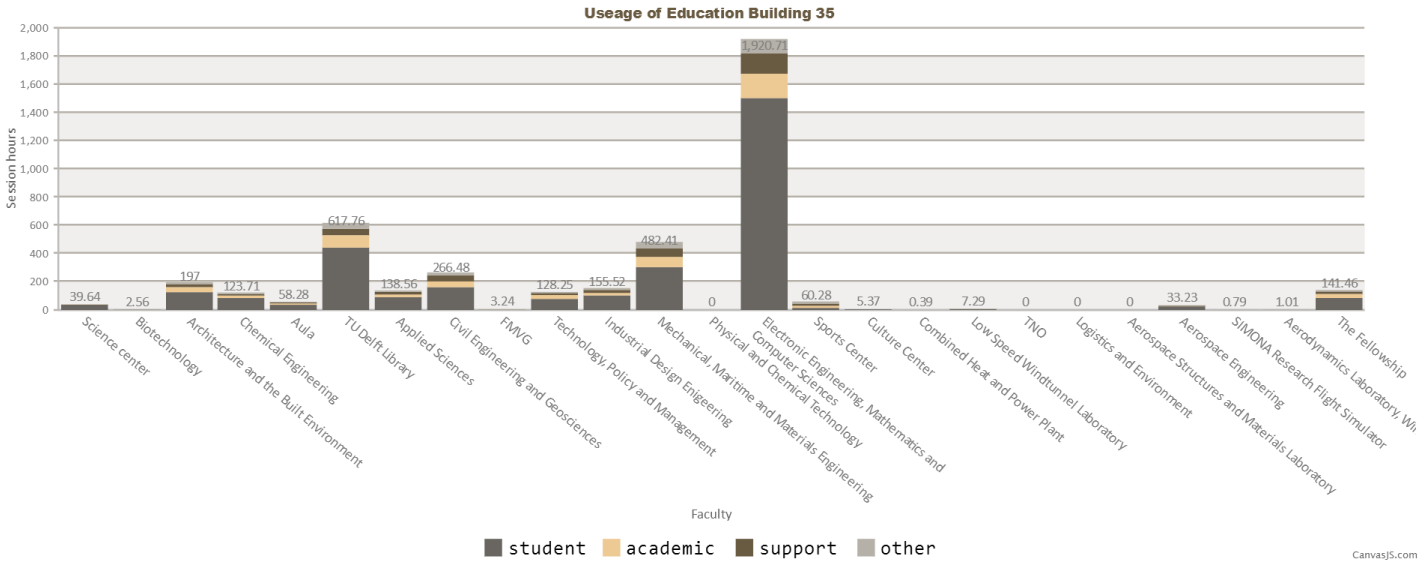


Figure A.39: Information of user group of building 35

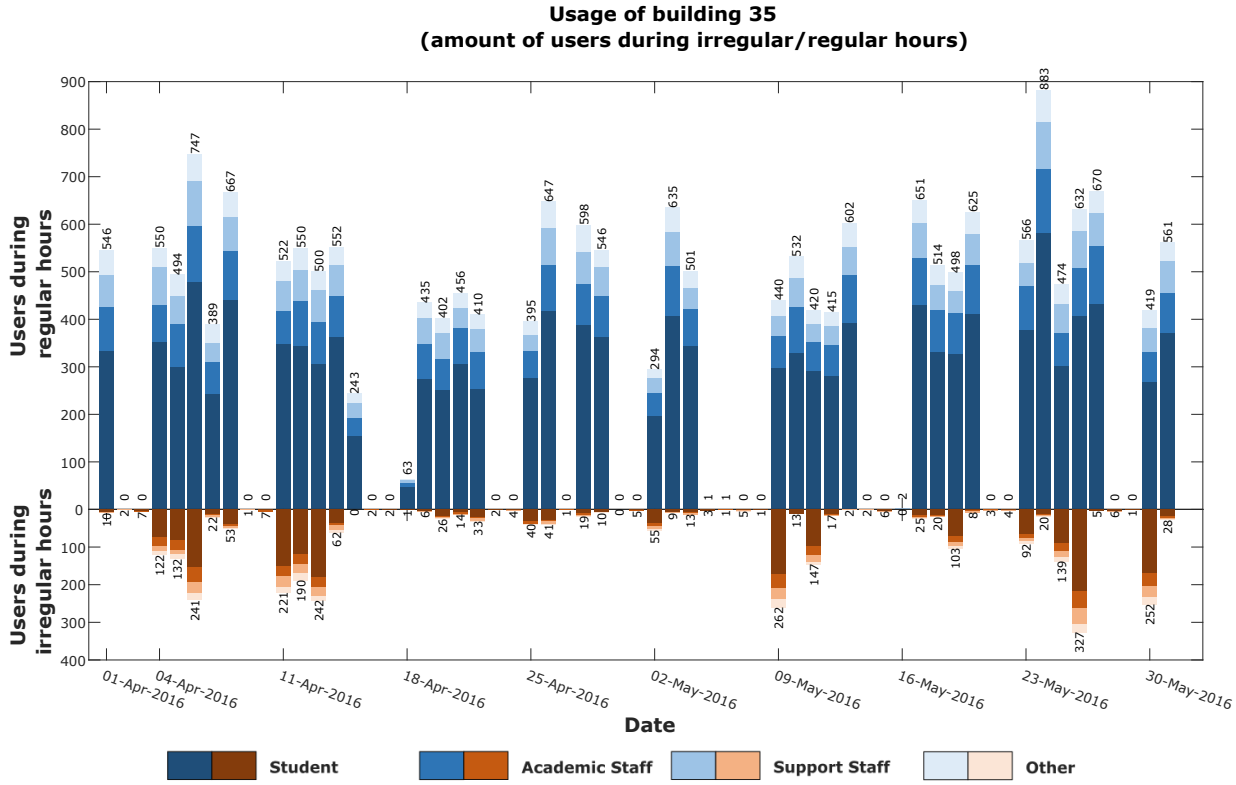


Figure A.40: Usage of building 35, amount of users present, divided by type of user

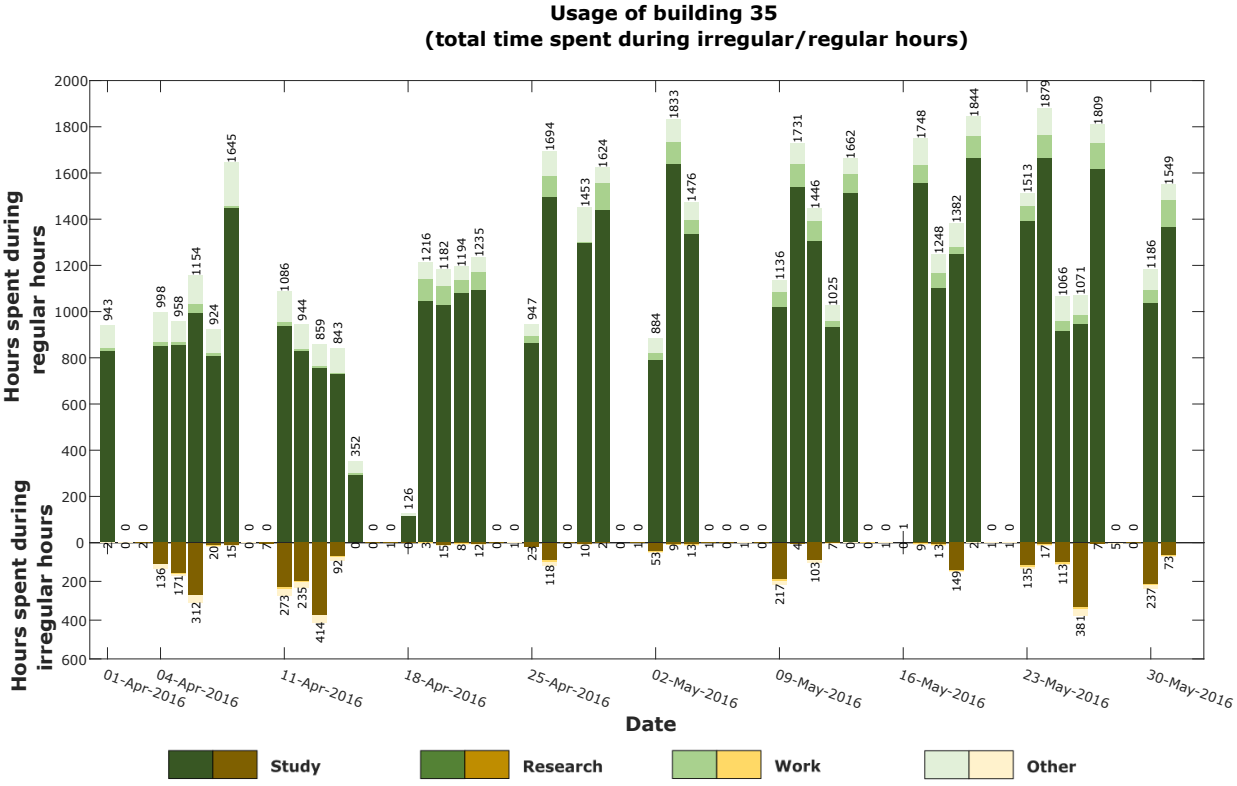


Figure A.41: Usage of building 35, the total time spent by users divided into activities

A.15. Building 36: Faculty of Electronic Engineering, Mathematics and Computer Sciences

Type of building: faculty building
Opening hours: Monday to Friday 06:30 - 22:30
Exceptions: weekends 06:30 - 23:00 with authorised campuscard

User number and time spent during irregular/regular hours
The Faculty of Electronic Engineering, Mathematics and Computer Sciences is more used after the exam week than during the exam. During weekdays in exam week, the daily user number is about 1950 in regular hours; each person spends about 4 hours per weekday in regular hours.
After the exam, the daily user number of the building during regular hours in weekdays increased to around 2300. The maximum occupancy of the building occurred on the second day of the 4th quarter which was 2969 people. Although the user number increases, the average time people spend in the building does not change.
During the irregular hours in weekdays, the average user number is about 75 per day; each user spends 1.5 hours. During the regular hours in weekend, user number increases to around 150-200 and the time each spends in the building also increases to 3.5 hours to 4 hours.
Main users of the building are students and the most common activity is study. Compared with the other educational buildings, the proportion of academic staff is slightly higher.

User group information
The Faculty of Engineering, Mathematics and Computer Science is mainly used by people from itself (in total 43,496 hours, on average 725 hours per days by all the users); therein, student (63.7%), academic staff (17.61%), support staff (10%) and others (8.63%). The building is barely used by people from other buildings; people from TU Delft library use the building for 32.8 hours per day on average respectively.

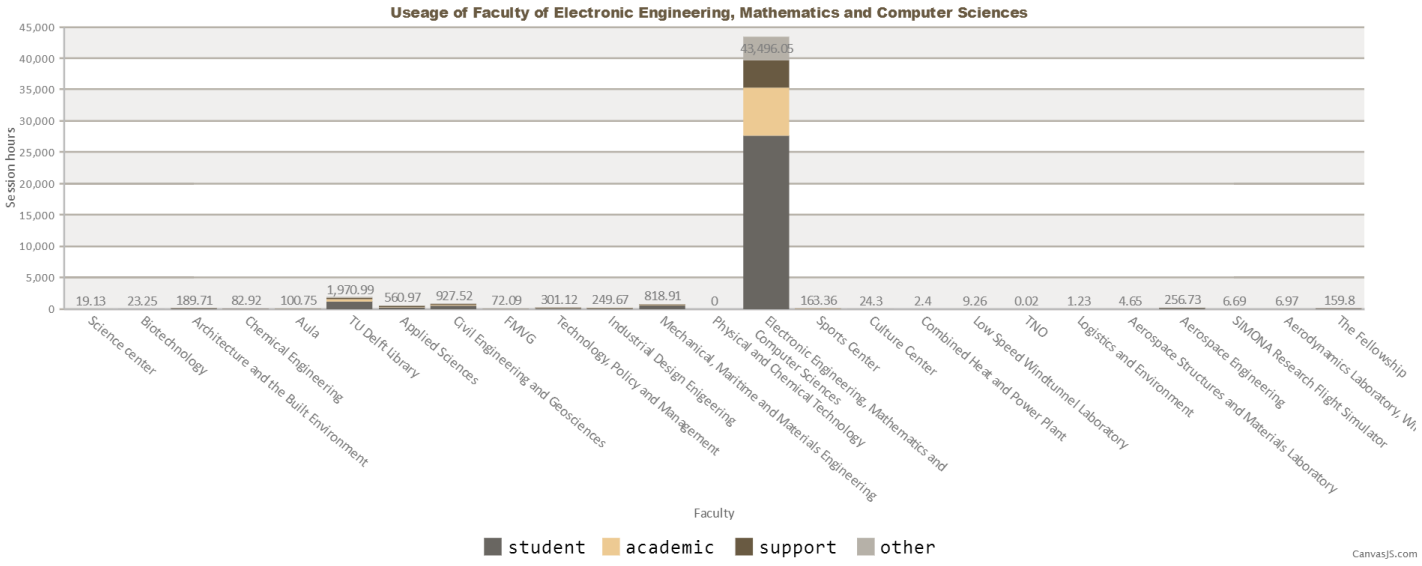


Figure A.42: Information of user group of building 36

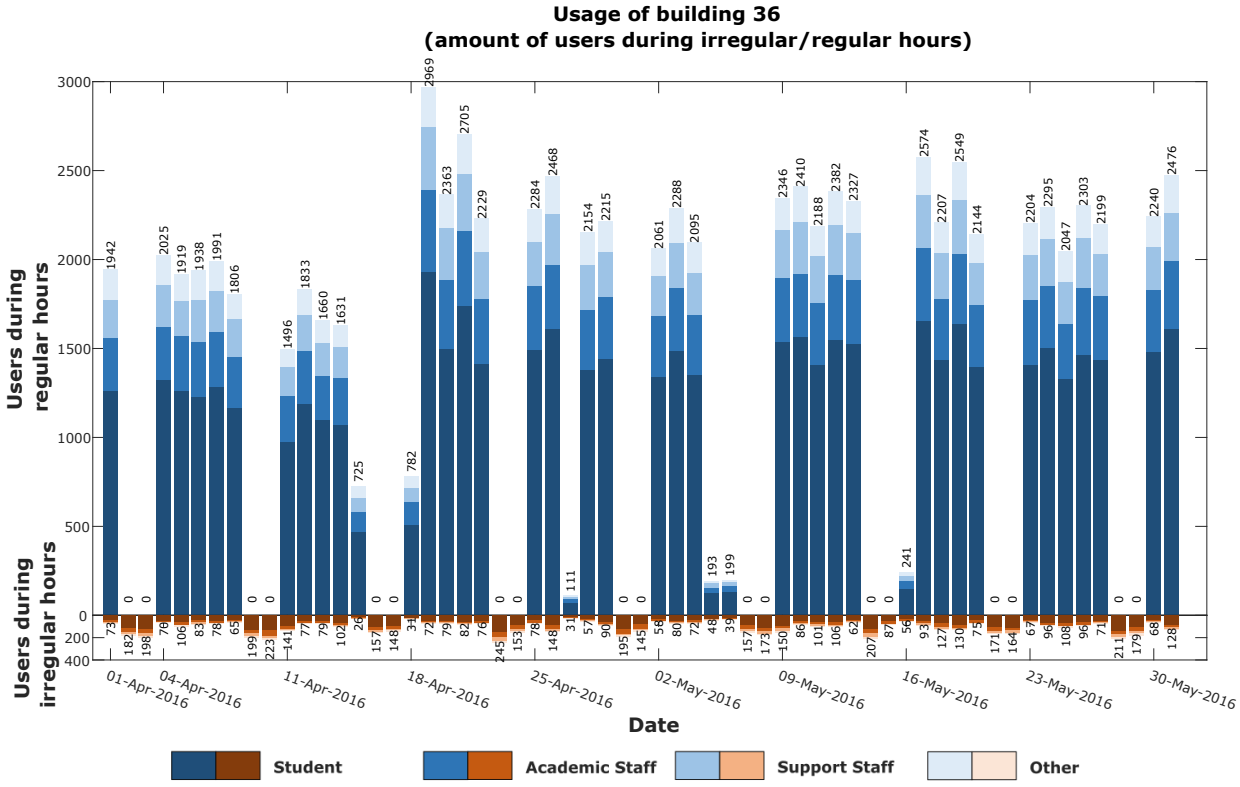


Figure A.43: Usage of building 36, amount of users present, divided by type of user

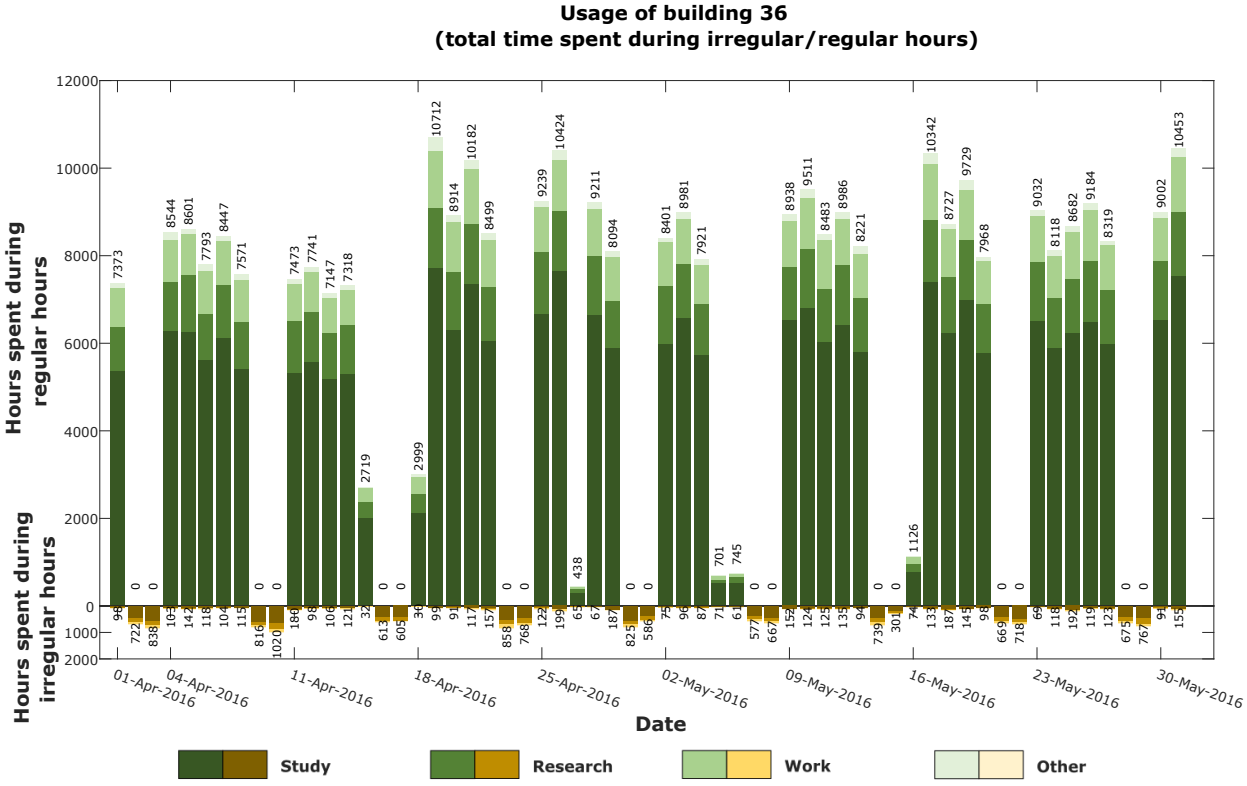


Figure A.44: Usage of building 36, the total time spent by users divided into activities

A.16. Building 37: Unit Sports

Type of building: sport and culture building
Opening hours: Monday to Friday 08:00 - 24:00 and weekend 08:00 - 21:00
Exceptions: Sportscafe is open until 01:00

User number and time spent during irregular/regular hours
The Unit Sports is more used during Monday to Thursday than during Friday to Sunday and is more used during the exam week than after the exam.
During the exam week, there are usually more than 800 users per weekday and more than 300 users per day in weekends. The maximum occupancy of the building occurred on 11th of April and the next day; 1184 and 1151 users used the building respectively in regular hours.
After the exam week, there are more than 700 users on Monday to Thursday, about 550 users on Friday and around 250-350 users on each day of the weekends.
Each user spends around 1.5 hours per day in Unit Sports on either weekdays or weekends.
The building is barely used during irregular hours. In weekdays, there are about 25 people per day use the building in irregular hours (half an hour per person) probably for preparing and cleaning up the space or in the sportscafe due to the extended opening time. In weekends, usually less than 10 people use the building per day.
The largest proportion of users is student; the most usual activity is undefined errand due to the insufficient category of activities. Compared with the other education buildings, more staff uses the building as well.

User group information
The Unit Sports is used by people from various buildings; people from itself use it for the longest time (in total 4,315 hours, on average 71.9 hours per days by all the users; therein, student (25%), academic staff (25%), support staff (25%) and others (25%); the data makes sense due to the relatively more staff working in the Unit Sports). The building is also frequently used by people from main faculties; in which people from Mechanical, Maritime and Material Engineering, TU Delft Library, Civil Engineering and Geosciences, Architecture and Built Environment, Mechanical, Maritime and Material Engineering, Electronic Engineering, Mathematics and Computer Sciences use the building for 10.7 hours, 9.85 hours, 9.7 hours and 8.6 hours per day by all the users on average. It can be indicated that people from different faculties go to the Unit Sports equally.

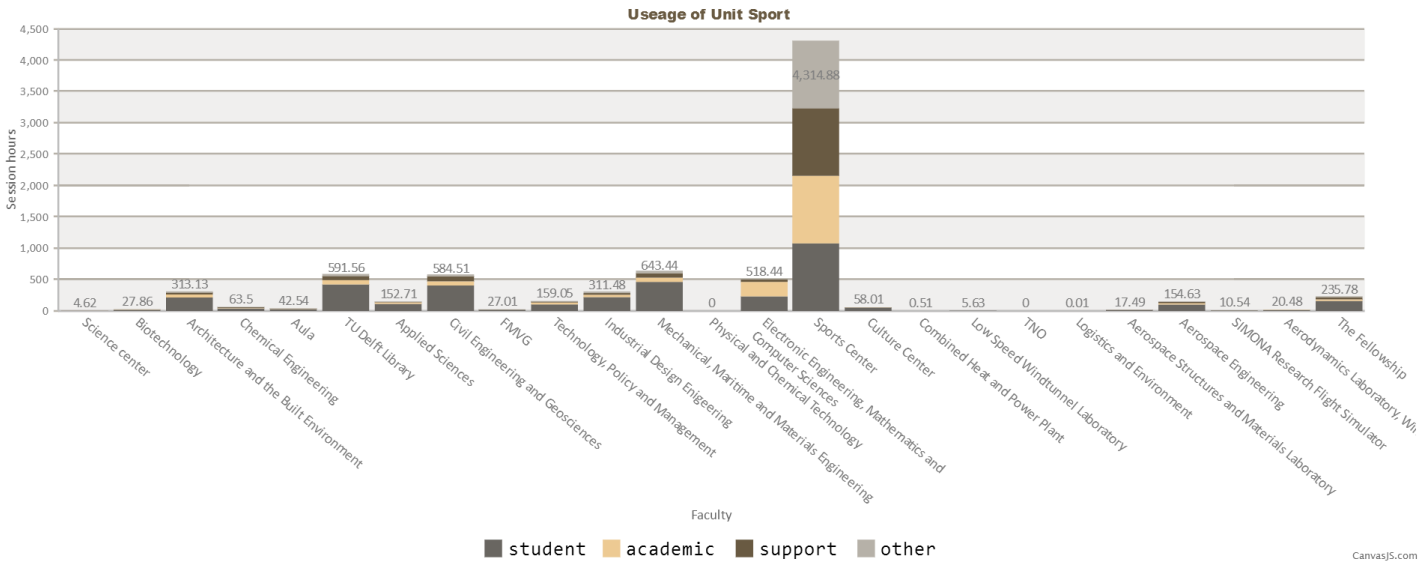


Figure A.45: Information of user group of building 37

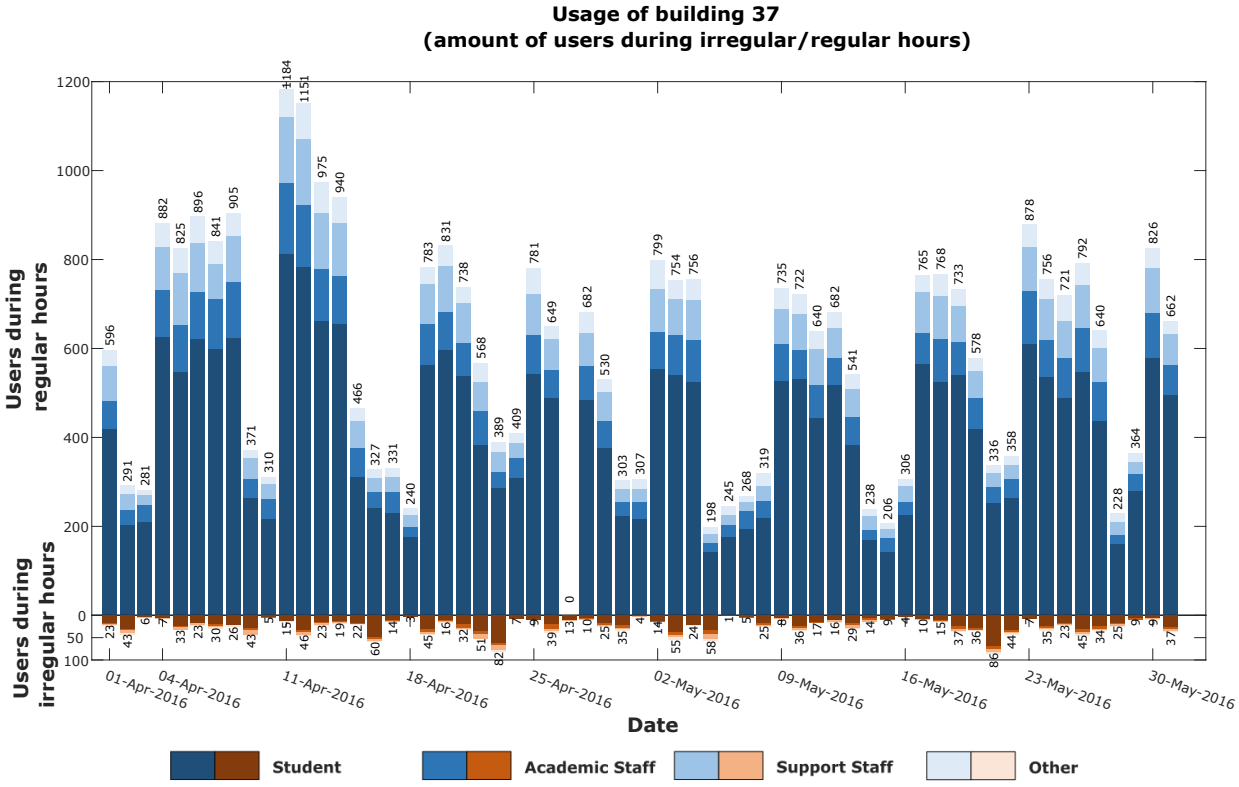


Figure A.46: Usage of building 37, amount of users present, divided by type of user

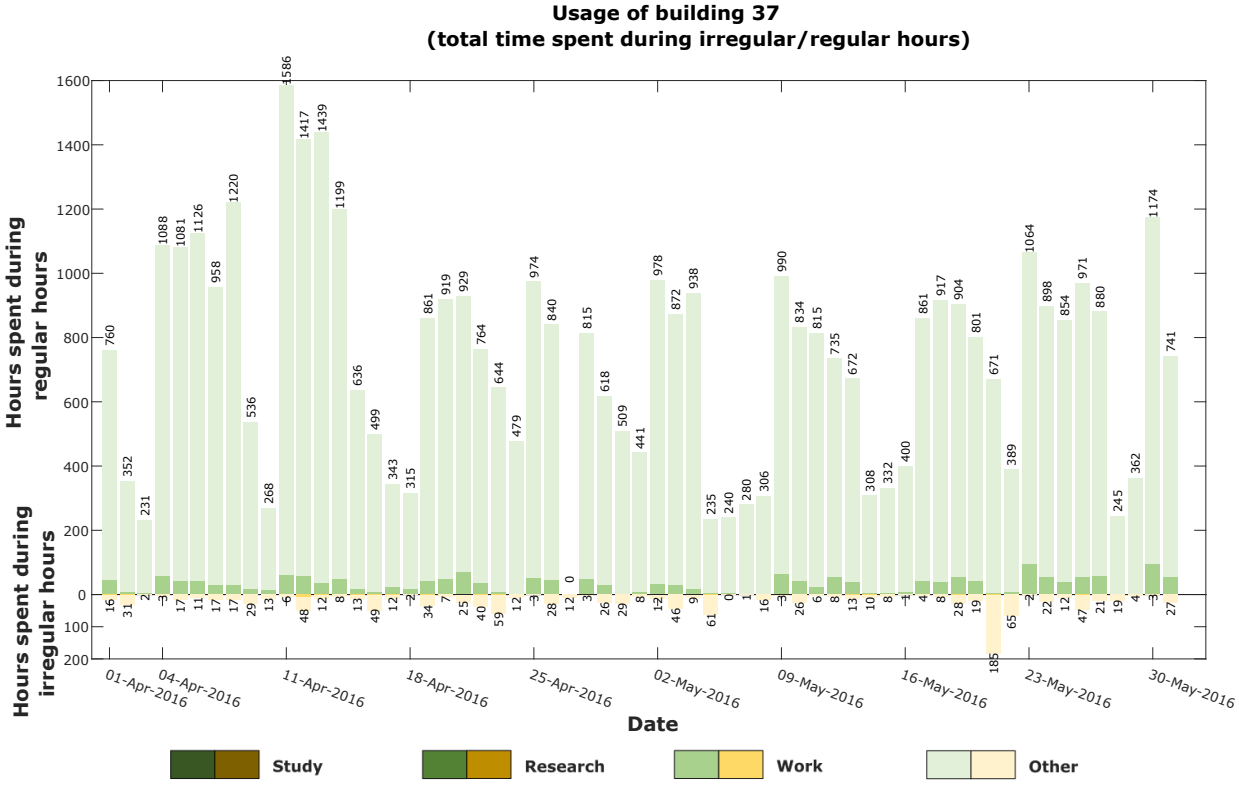


Figure A.47: Usage of building 37, the total time spent by users divided into activities

A.17. Building 38: Unit Culture

Type of building: sport and culture building
Opening hours: Monday to Friday 08:00 - 24:00, Saturday 10:00 - 17:00, and Sunday 14:00 - 24:00
Exceptions: none

User number and time spent during irregular/regular hours
The Unit Culture is frequently used both during weekdays and weekends in regular hours compared with the other education buildings.
During the exam week, in regular hours in weekdays, everyday user number varies from 193 to 348 but the average daily user count is around 220 people; the average time each user spends in Unit Culture is about 1.5 hours per weekday. After the exam, the daily user number of Unit Culture varies from 94 to 512 but the average number is 350 to 450 per weekday; each person spends around 1.5 hours per day in there.
During the regular hours in weekends, usually there are about 100 people on each Saturday and Sunday, each person spends 1 hour to 1.5 hours there.
The building is not frequently used during irregular hours in weekdays; however, on every Saturday and Sunday, there are minimum 31 (maximum 139) people using the building for 1.5 hours to 3 hours.
The largest proportion of users is student; the most usual activity is undefined errand due to the insufficient category of activities.

User group information
The Unit Culture is used by people from various buildings; people who go to the Unit Sports usually use it for the longest time (in total 863 hours, on average 14.4 hours per days by all the users; therein, student (25%), academic staff (25%), support staff (25%) and others (25%)). The Unit Culture is also frequently used by people from main faculties; in which people from Electronic Engineering, Mathematics and Computer Sciences, Civil Engineering and Geosciences, Industrial Design Engineering TU Delft Library use the building for 5.1 hours, 4.6 hours, 4.4 hours and 3.46 hours per day by all the users on average. It can be indicated that compared with the Unit Sports, students go to culture more often than staff.

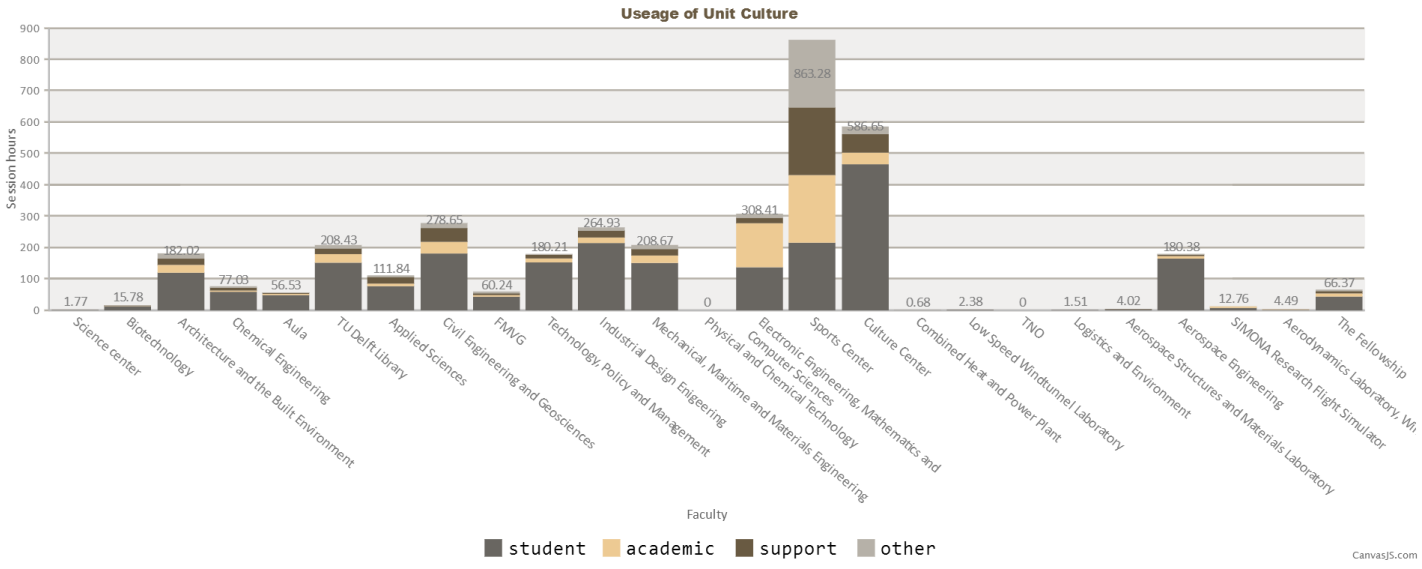


Figure A.48: Information of user group of building 38

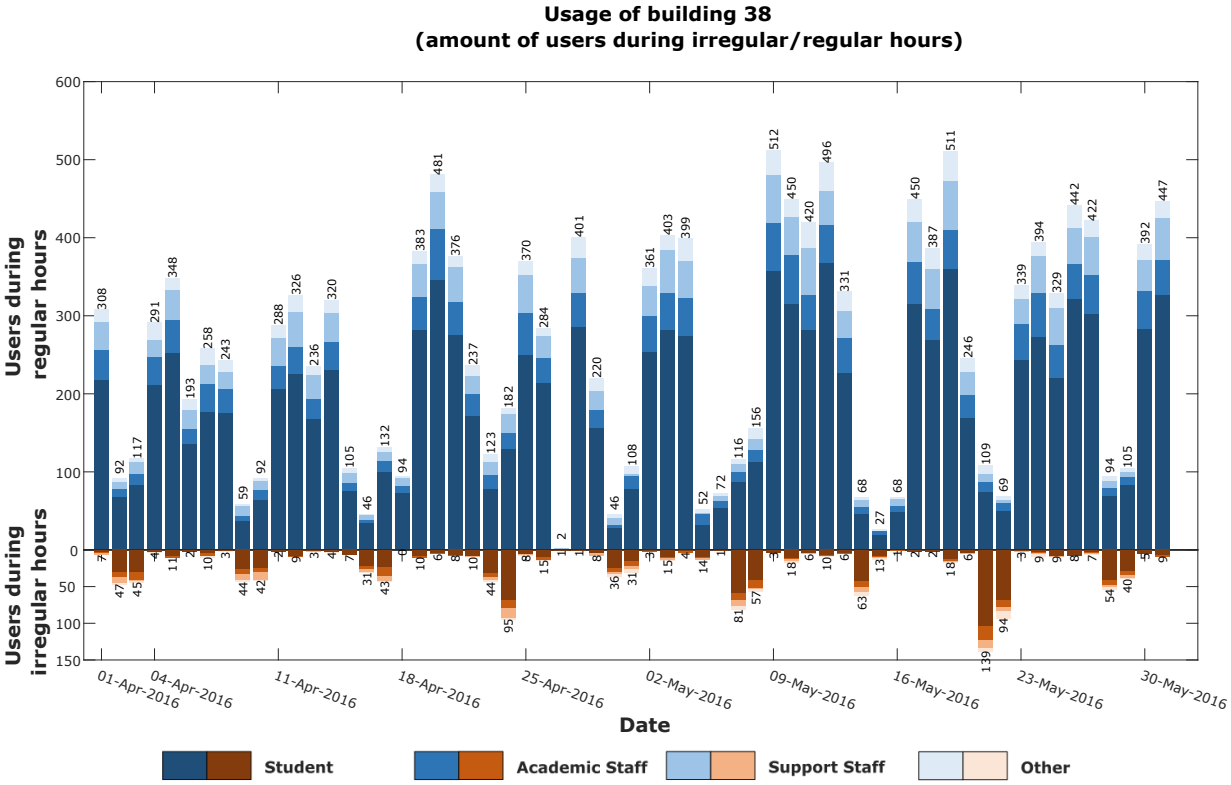


Figure A.49: Usage of building 38, amount of users present, divided by type of user

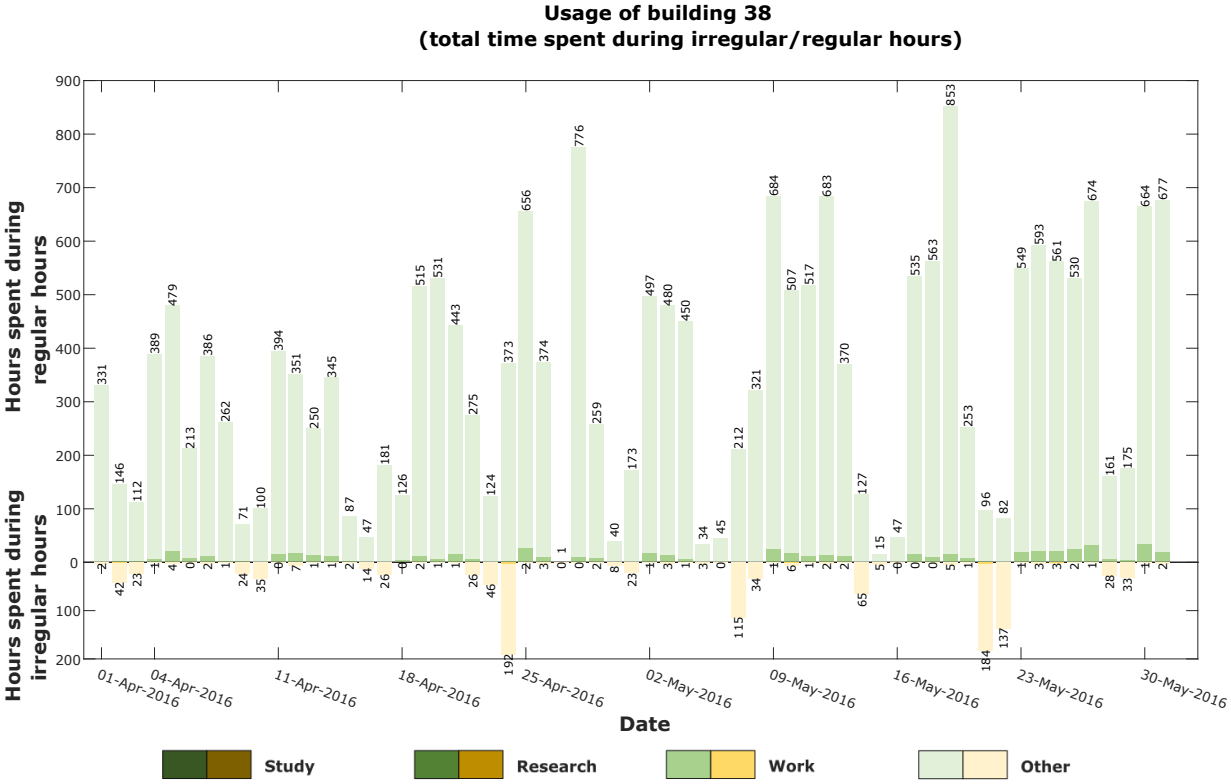


Figure A.50: Usage of building 38, the total time spent by users divided into activities

A.18. Building 43: Combined Heat and Power Plant

Type of building: research building
Opening hours: not specified
Exceptions: none

User number and time spent during irregular/regular hours
The reason why all the records are in irregular hours is that the opening time of the building is not specified.
During the exam week, there are around 15 users per weekday use the building, most of them are academic staff. Each user spends around 3.5 hours per weekday in the building for undefined errands. The building is barely used during weekends.
After the exam week, the user number slightly increases to around 18. The maximum occupancy of the building occurred on 5th May 2016; in total 51 users were in the building during that day but the average time (around half an hour per person) each person spent there is less than what (around 2 hours to 3 hours per person) on other weekdays.
On 31th May 2016, the building was used for the longest time (from 1st of April to 31st of May) which is 85 hours (2.8 hours per person).

User group information
The building of Combined Heat and Power Plant is mainly used by people from itself (in total 323 hours, on average 5.4 hours per days by all the users); therein, student (11.56%), academic staff (85.9%), support staff (2.53%) and others (0.02%). The building is barely used by people from other buildings. It can be indicated that the building is seldom used by students; most of the users are academic staff which is consistent with the fact that the building is a research building.

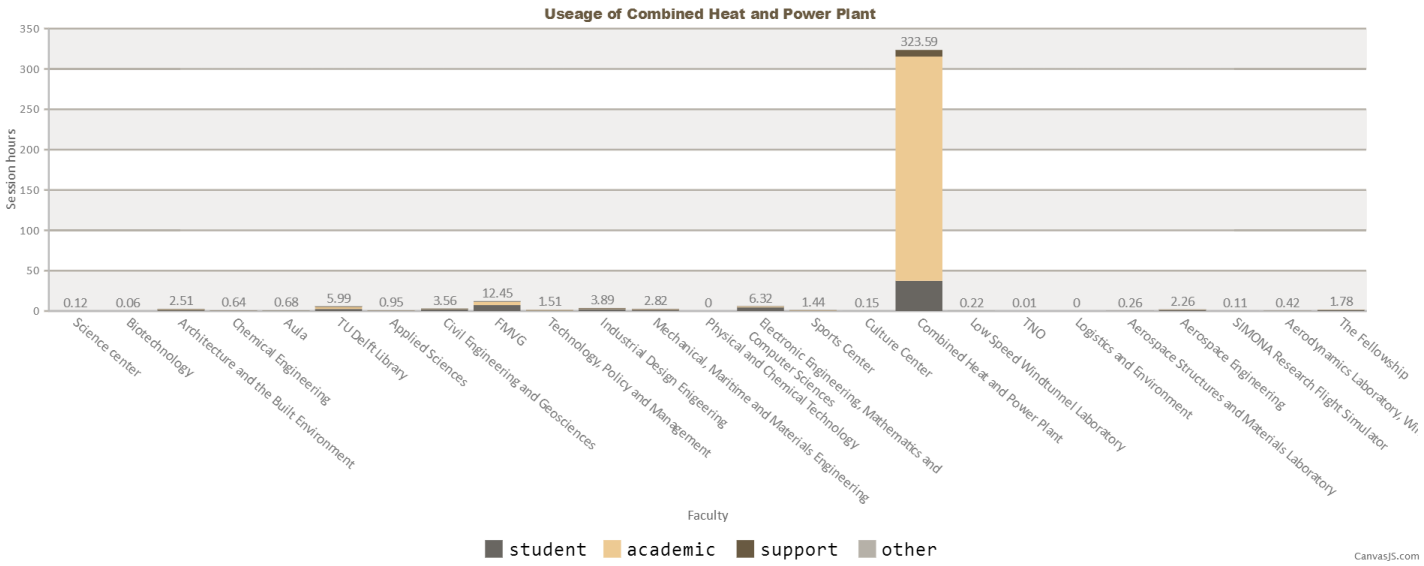


Figure A.51: Information of user group of building 43

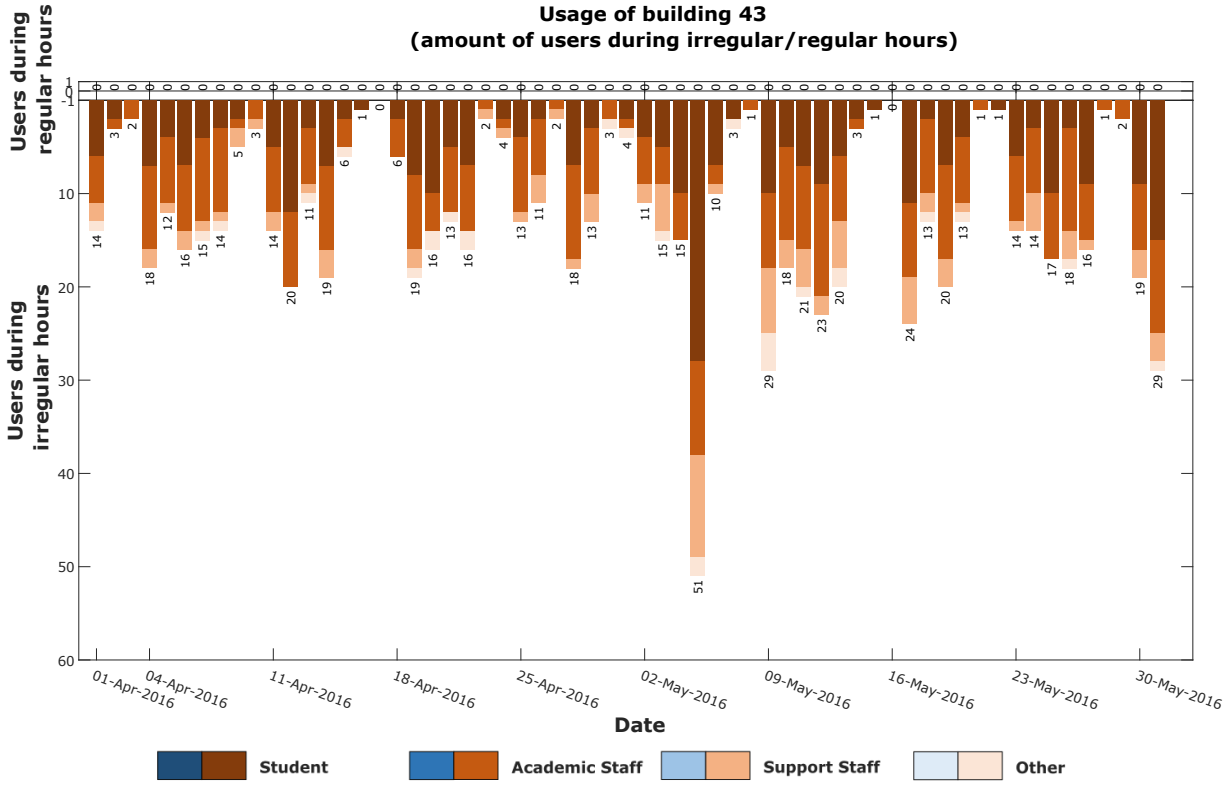


Figure A.52: Usage of building 43, amount of users present, divided by type of user

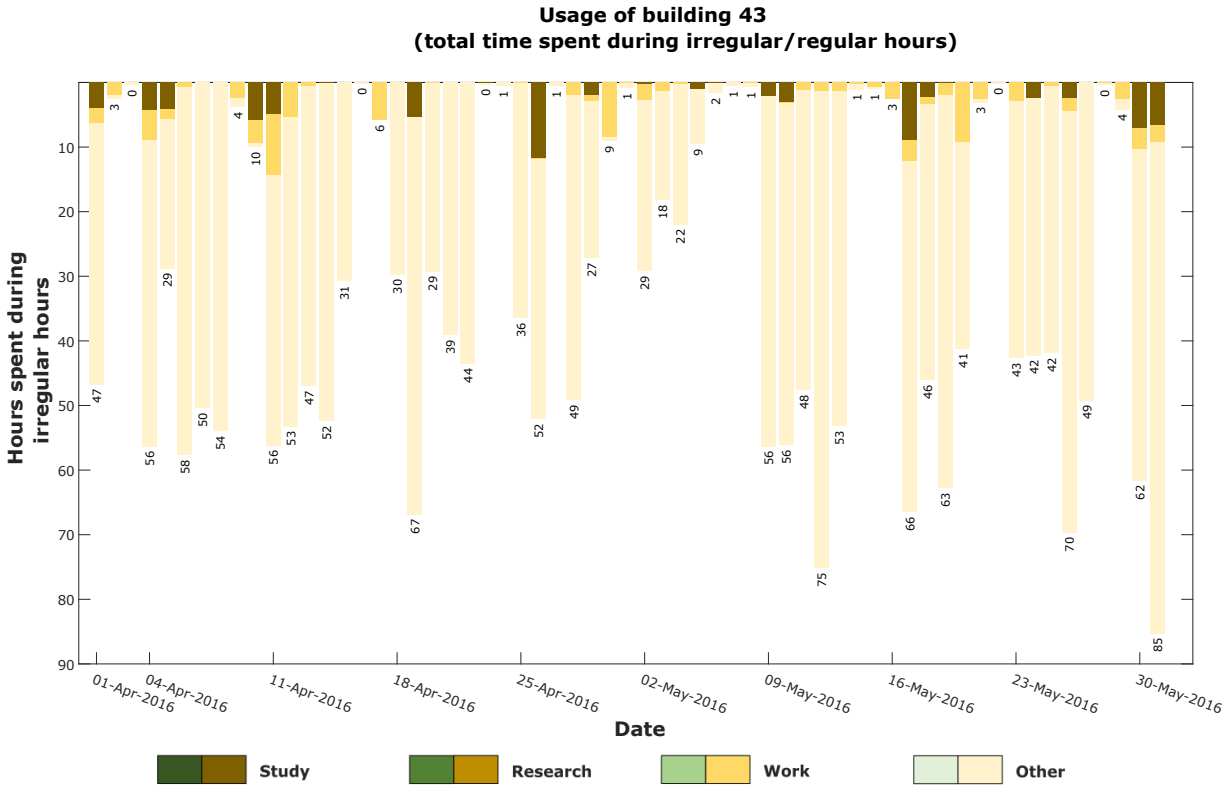


Figure A.53: Usage of building 43, the total time spent by users divided into activities

A.19. Building 45: Low Speed Windtunnel Laboratory

Type of building: research building
Opening hours: Monday to Friday 08:00 - 18:00
Exceptions: none

User number and time spent during irregular/regular hours
During the exam week, on average, there are 40 to 45 users in the building in regular hours; each user spends around 2.5 hour to 3 hours on average in the building per day. The maximum occupancy of the building occurred on the second day of the 4th quarter (19th of April); 117 people were in the building during regular hours; this is possibly due to most of the courses started on that day. In weekdays after the exam week, there are on average 70 people per day in the building in regular hours; each user spends around 1.5 hours to 2 hours per day. An obvious decrease of user number can be observed on every Fridays.
In irregular hours in weekdays during the exam week, there are on average 20 users per day in the building; each person spends around 2 hours there. In weekends, people spend less time there (around 1 to 1.5 hours per day). After the exam week, there is a slight increase of user number during irregular hours in weekdays; however, the time they spend also decreases (around 1 to 1.5 hours per day).

User group information
The building of Low Speed Wind-tunnel Laboratory is mainly used by people from itself (in total 734 hours, on average 12.2 hours per days by all the users); therein, student (77%), academic staff (7.3%), support staff (5.8%) and others (10%). The building is barely used by people from other buildings; in which people from TU Delft library use the building for less than one hours per day by all the users on average.

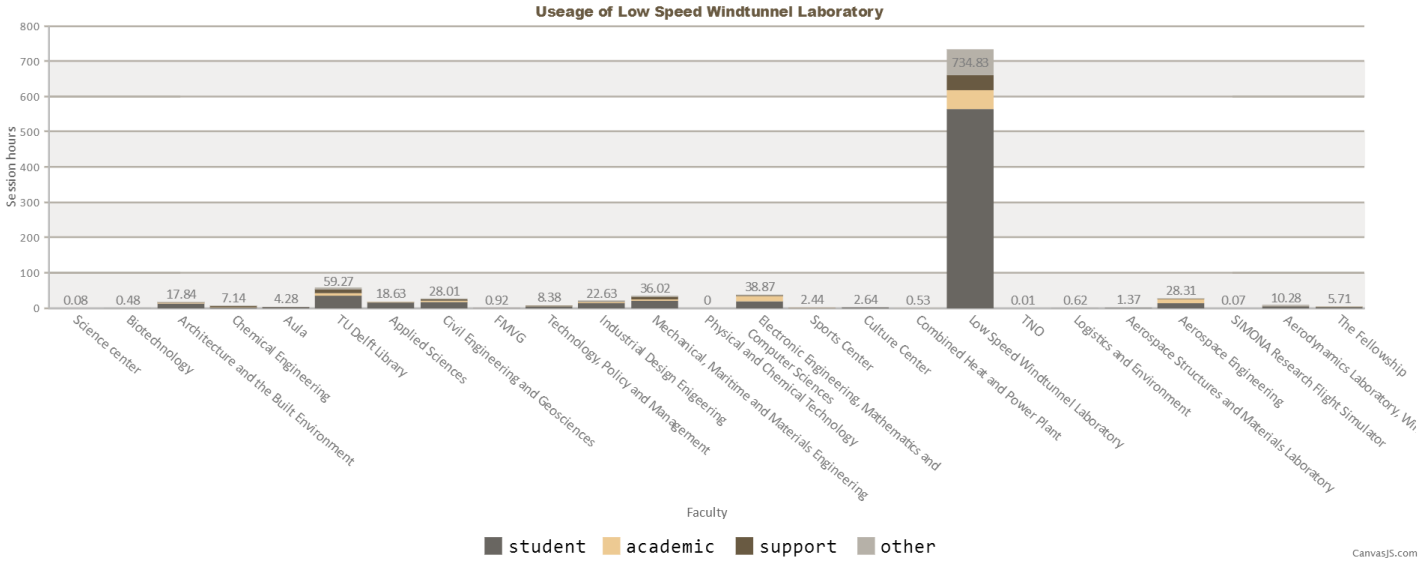


Figure A.54: Information of user group of building 45

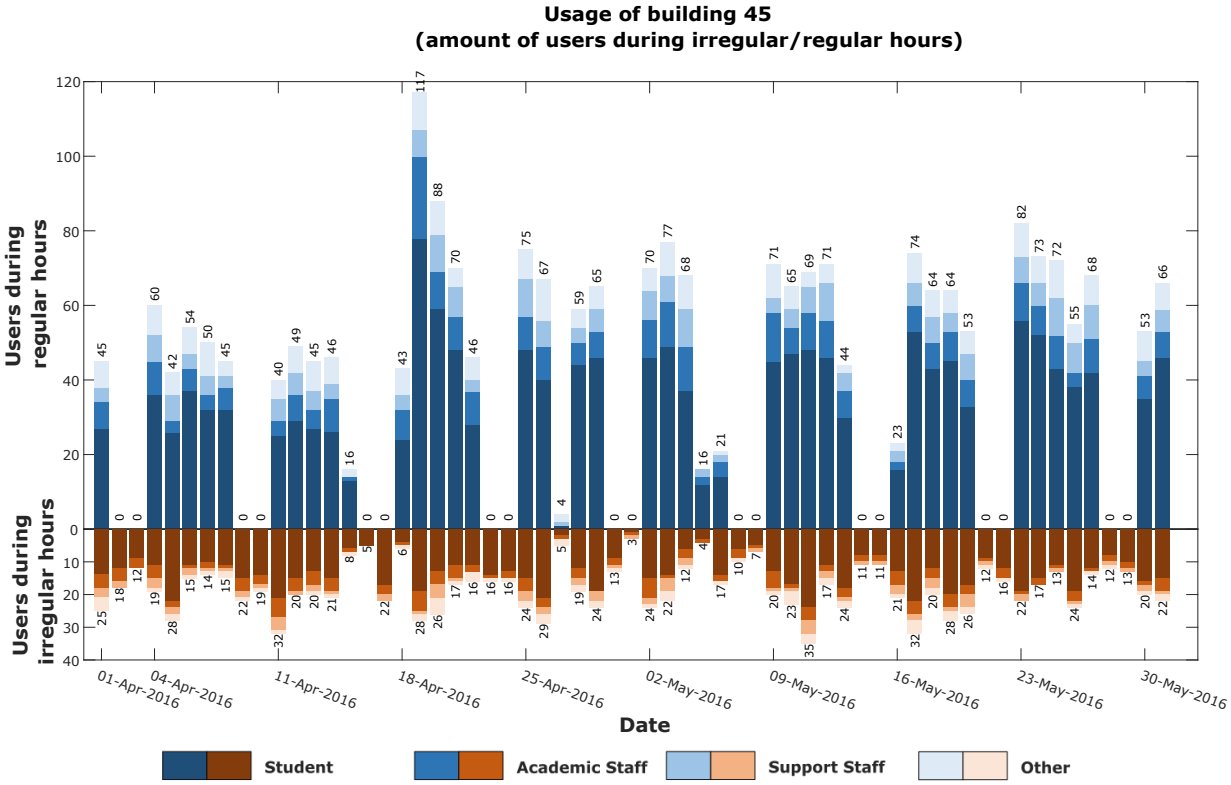


Figure A.55: Usage of building 45, amount of users present, divided by type of user

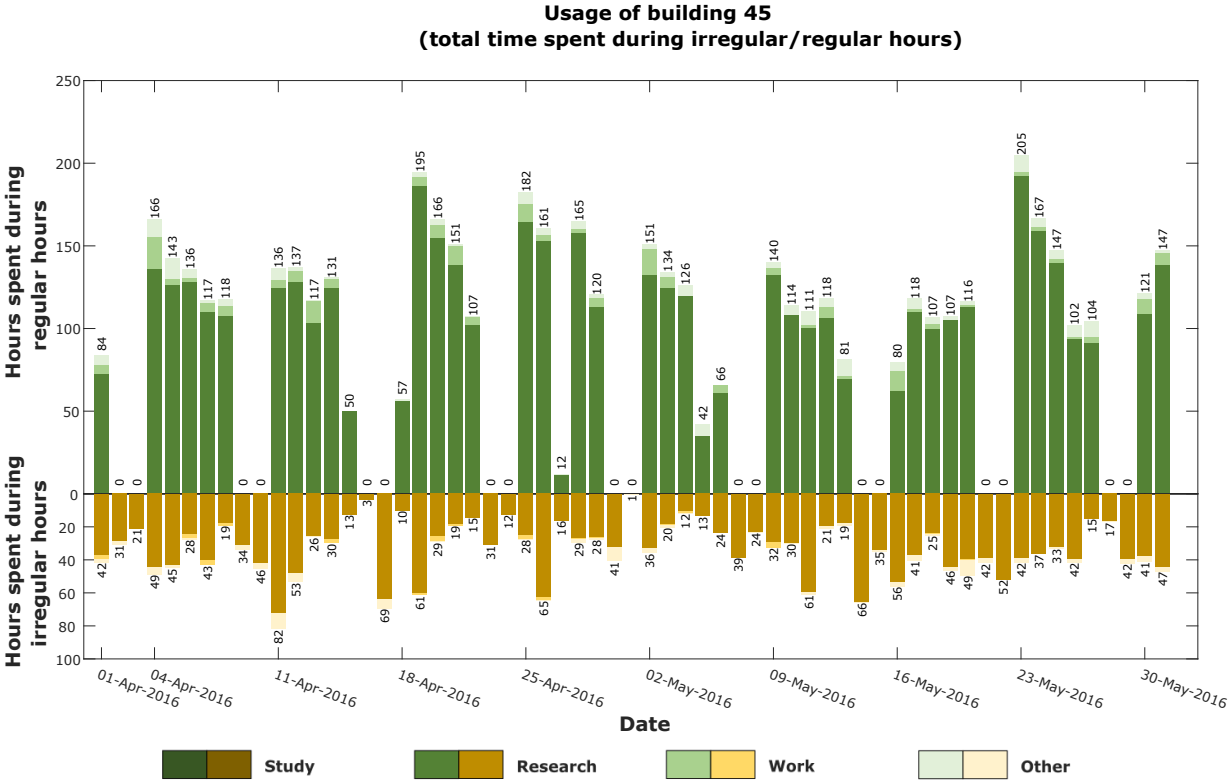


Figure A.56: Usage of building 45, the total time spent by users divided into activities

A.20. Building 46: TNO

Type of building: research building
Opening hours: Monday to Friday 08:00 - 18:00
Exceptions: none

Users number and time spent during irregular/regular hours
The maximum user number in one weekday is nine; each person spent less than 15 minutes in TNO. On average, during regular hours, one user spends less than half an hour in the building. What is more, it is almost the same for irregular hours. It can be indicated that the building is left unused or barely used. The records in the allocated in the database could be caused by people passing the building.

User group information
The TNO is used by people from various buildings; people from Electronic Engineering, Mathematics and Computer Sciences use it for the longest time (in total 5.42 hours, on average less than one hour per days by all the users; therein, student (67%), academic staff (16%), support staff (10%) and others (6.5%)). TNO is also frequently used by people from main faculties; however, the total usage of the building is relatively low compared with the other buildings; the session durations are probably just from people passing by the building. It can be indicated that the building is left unused from 1th April 2016 to 31th May 2016.

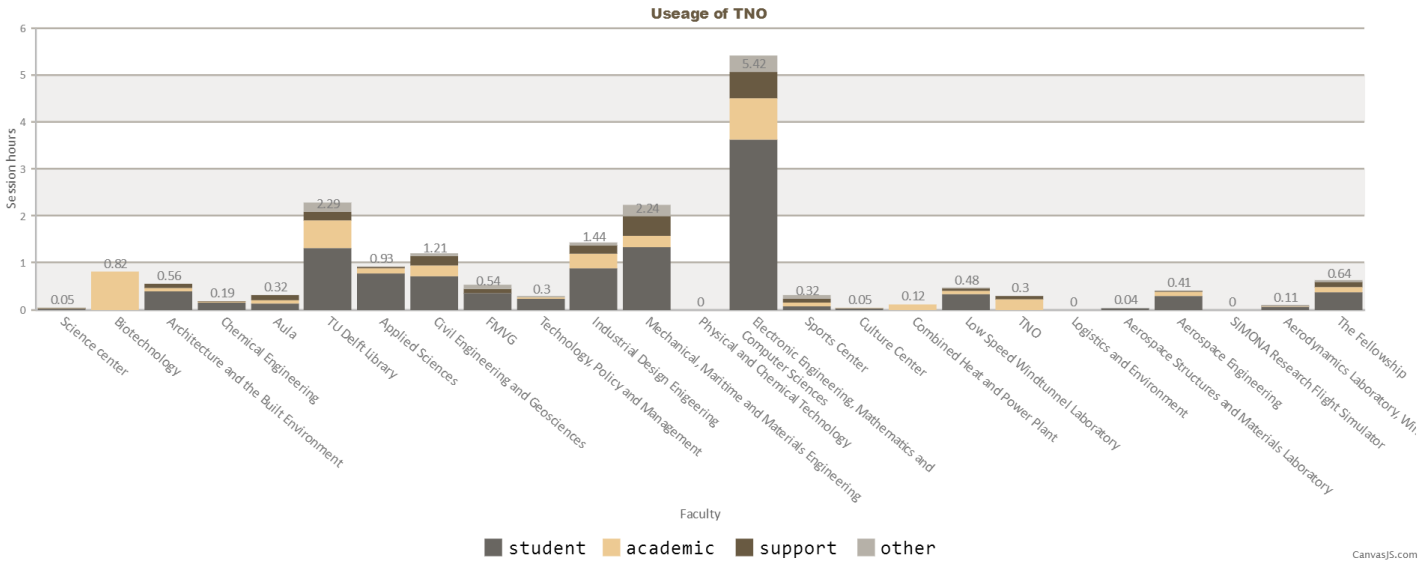


Figure A.57: Information of user group of building 46

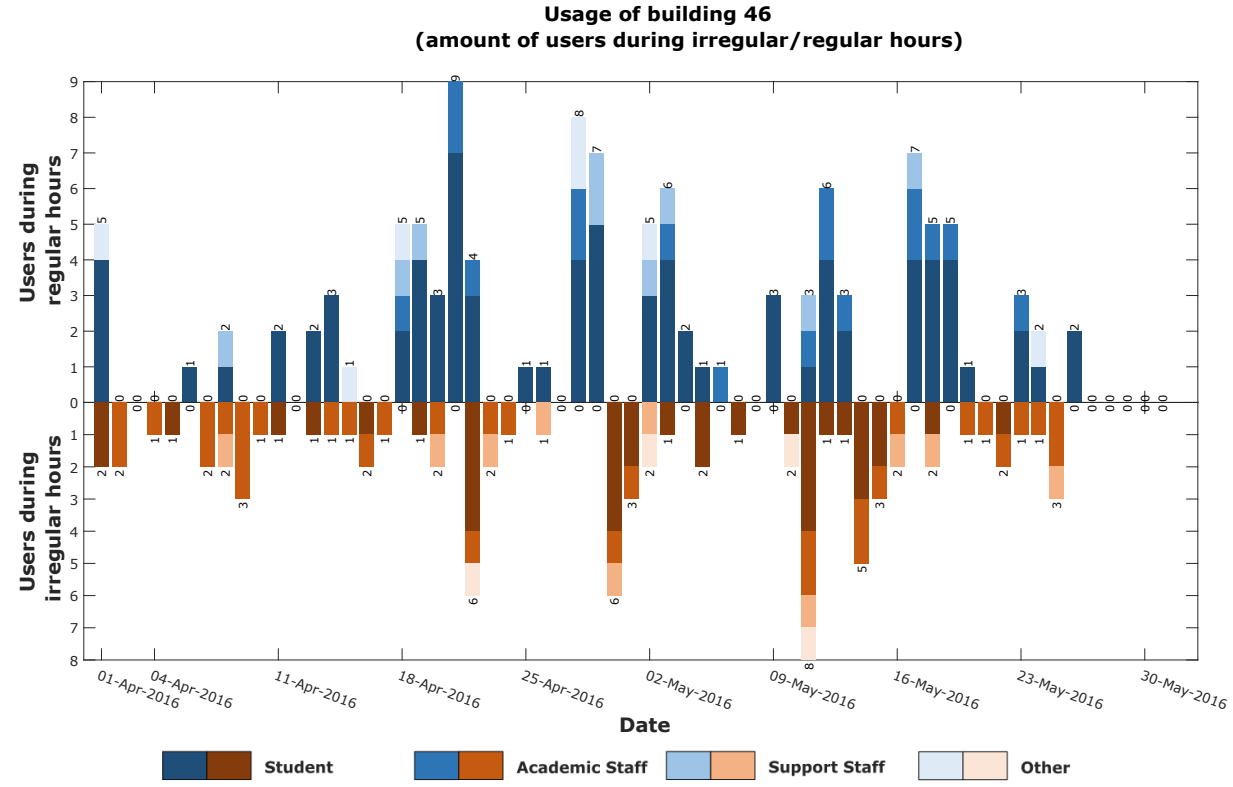


Figure A.58: Usage of building 46, amount of users present, divided by type of user

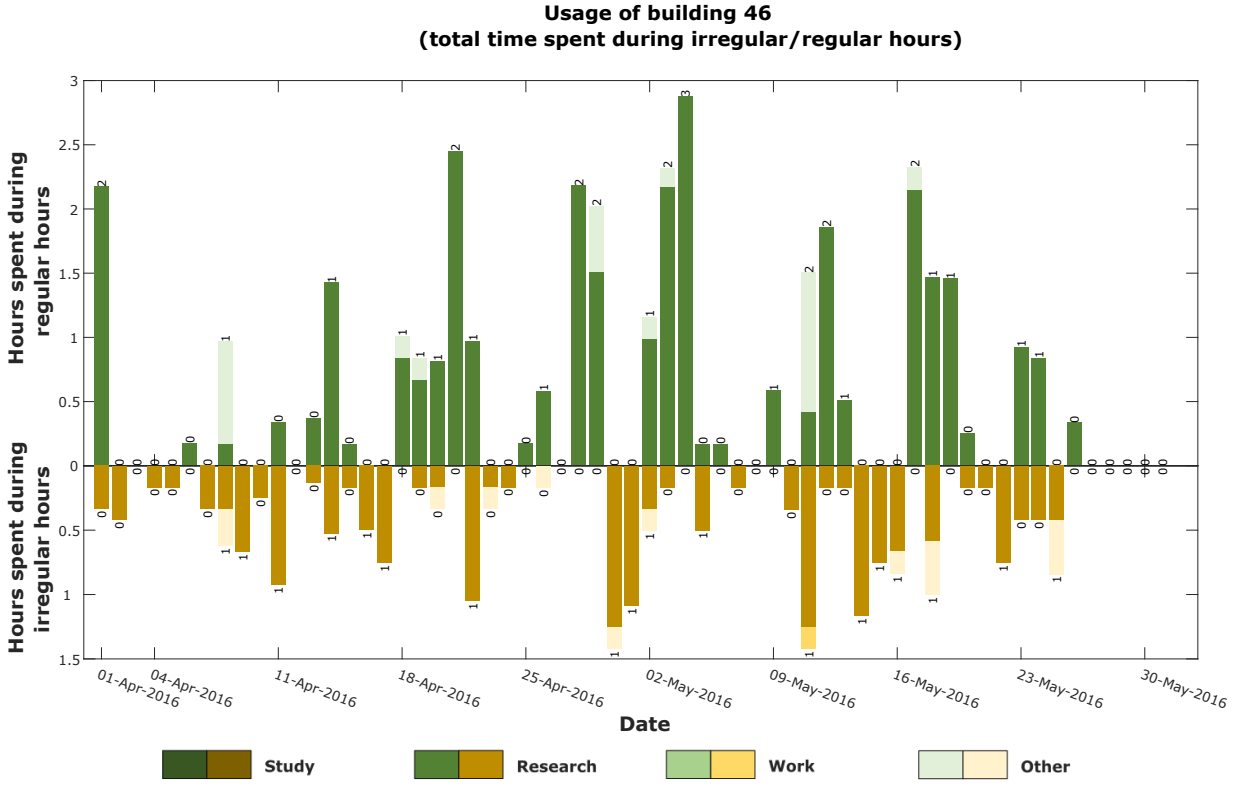


Figure A.59: Usage of building 46, the total time spent by users divided into activities

A.21. Building 60: Logistics and Environment

Type of building: facility management and real estate building
Opening hours: Monday to Friday 07:30 - 16:00
Exceptions: none

User number and time spent during irregular/regular hours
It can be found that almost for every weekday, there are a quite amount of users (about half of number of the daily users in regular hours) in irregular hours (around 13 users during exam week, around 20 users after the exam). It could be caused by the relative early closing time therefore people prefer to stay longer for finish their work. There seldom are users in the building in weekends either during regular or irregular hours.
Although students are the largest proportion of the user group; the most frequent usage of the building is other or working. In regular hours, it is common that one user spend one hour in working and two hours in doing other things per weekday. In irregular hours, users usually spend less than half an hour in other things or working.

User group information
The building of Logistics and Environment is mainly used by people from itself (in total 452 hours, on average 7.53 hours per days by all the users); therein, student (68%), academic staff (0%), supports staff (17.75%) and others (14%). The building is barely used by people from other buildings; people from FMRE use the building for 10 minutes per day on average for may be just walking by it.

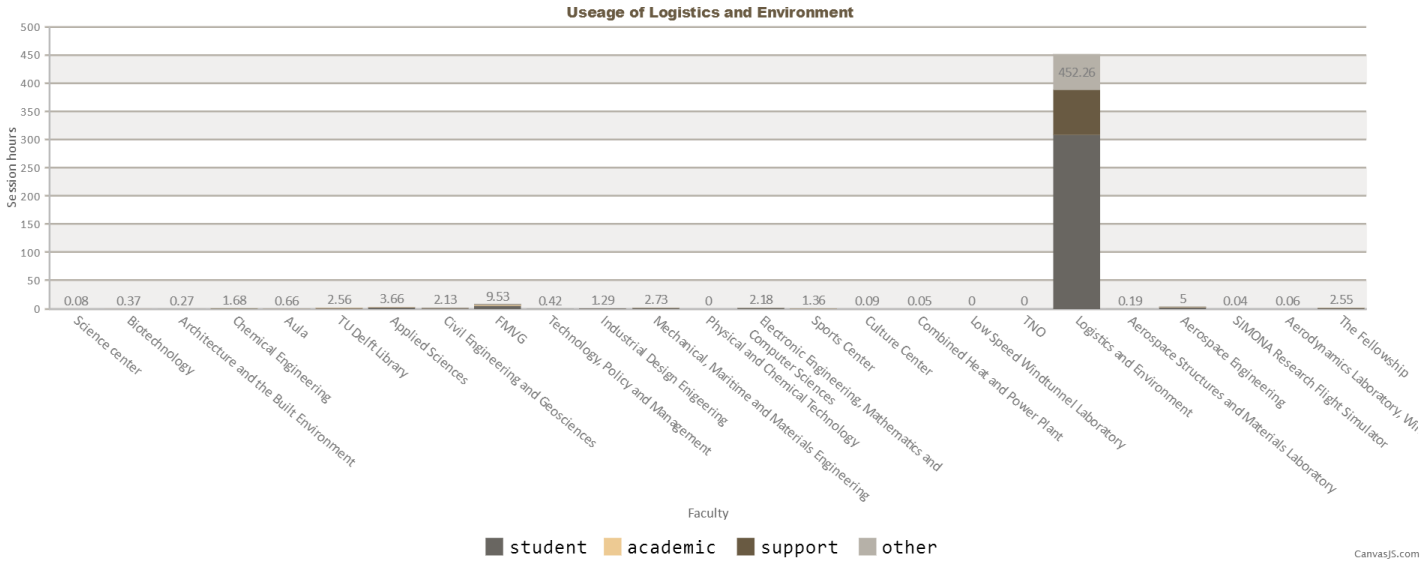


Figure A.60: Information of user group of building 60

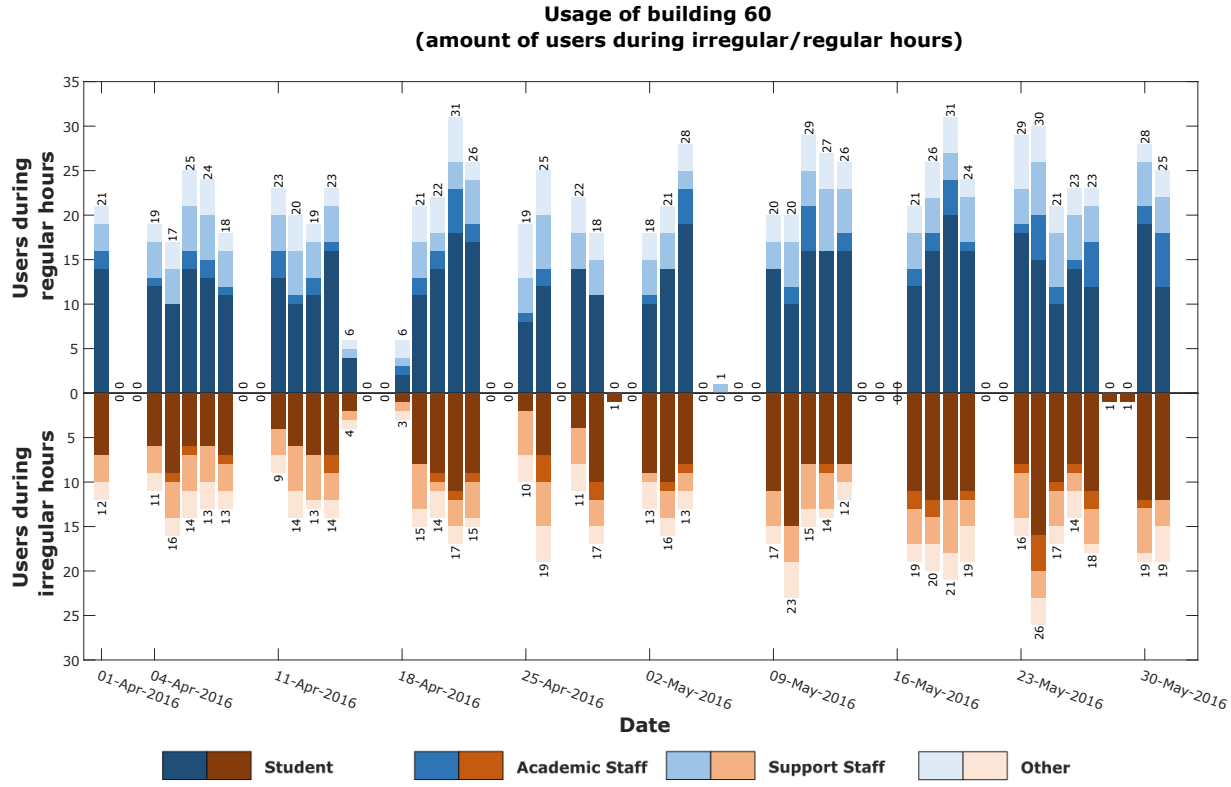


Figure A.61: Usage of building 60, amount of users present, divided by type of user

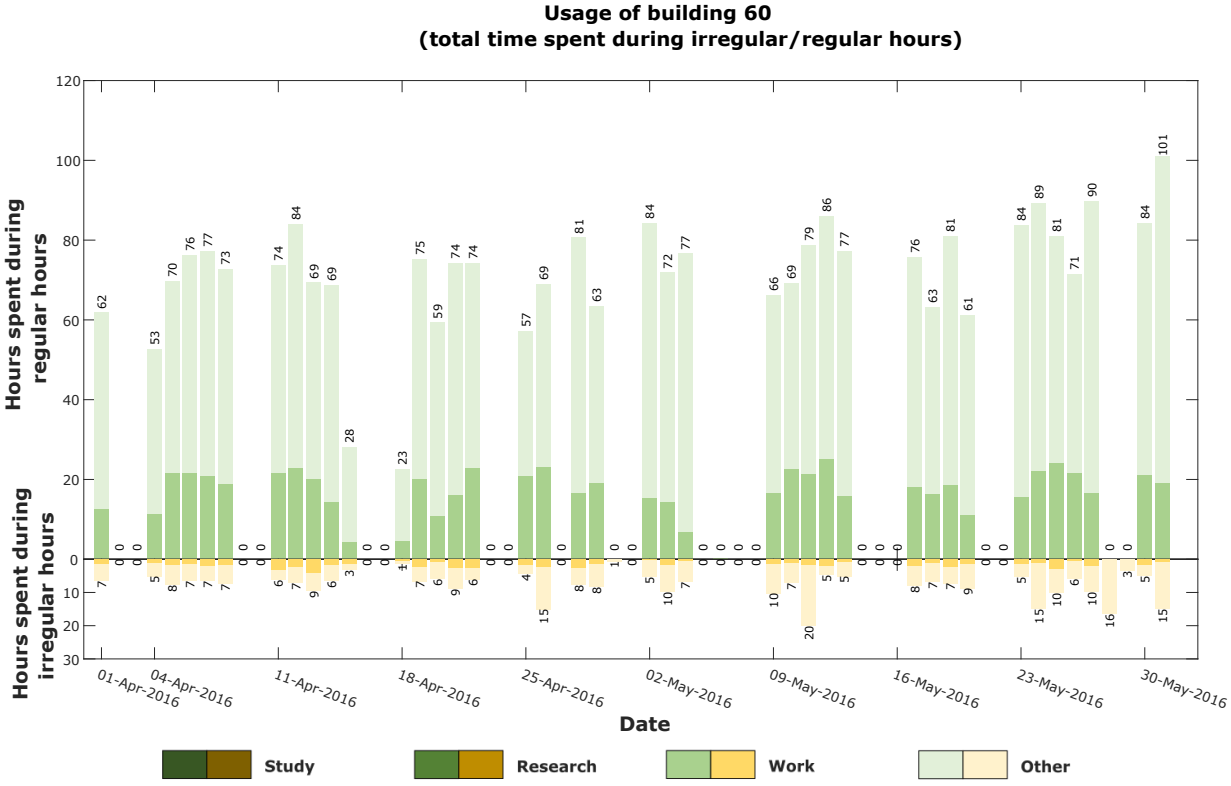


Figure A.62: Usage of building 60, the total time spent by users divided into activities

A.22. Building 61: Delft Aerospace Structures and Material Laboratory

Type of building: research building
Opening hours: Monday to Friday 07:00 - 22:00
Exceptions: none

Users and time spent during irregular/regular hours
During exam week, more users were using the Aerospace Structures and Material Laboratory in weekends compared with the weekends after exam week. The peak of using this building in weekends was found to be 73 users, on 10th of April (Sunday). After the exam week, usually there were 10 to 20 users on Saturdays and 0 to 4 users on Sundays; each user spent half to one hour per day. People seldom work after 10 pm during weekdays, however, on average; there will be one to two people work after 10 pm for one to two hours. On Fridays, usually there would be a decrement in user number. There were on average 30 to 50 more users using the building in May than in April. The time per user per day spent in the laboratory during regular time is 2.5 hours to 3.5 hours which is relatively longer compared with what during irregular time.
The daily user number during weekdays is around 100 to 150 during exam week and 150 to 200 after the exam. The number drops about 50 on Fridays. During regular hours of weekdays, people usually work for 2.5 hour to 3.5 hours per day; during weekends, people usually work half to one hour per day. What is more, more users work on Saturdays than on Sundays after the exam which is the opposite during the exam week. During the exam week, more users work in weekends than after the exam.

User group information
The Aerospace Structure and Materials Laboratory is mainly used by people from Aerospace Engineering (in total 1,343 hours, on average 22.4 hours per days by all the users; therein, student (93%), academic staff (2%), support staff (3.2%) and others (1%)), itself (in total 1,079 hours, on average 18 hours per days by all the users; therein, student (60%), academic staff (10.9%), support staff (7.8%) and others (21%)) and from the Fellowship (in total 217 hours, on average 3.6 hours per days by all the users; therein, student (71%), academic staff (11%), support staff (10%) and others (6%)). People from SIMONA Research Flight Simulator Laboratory, TU Delft Library also use the Laboratory. Most of these buildings are aerospace related which makes sense; the Fellowship building has better computer which supports the computation of experiments.

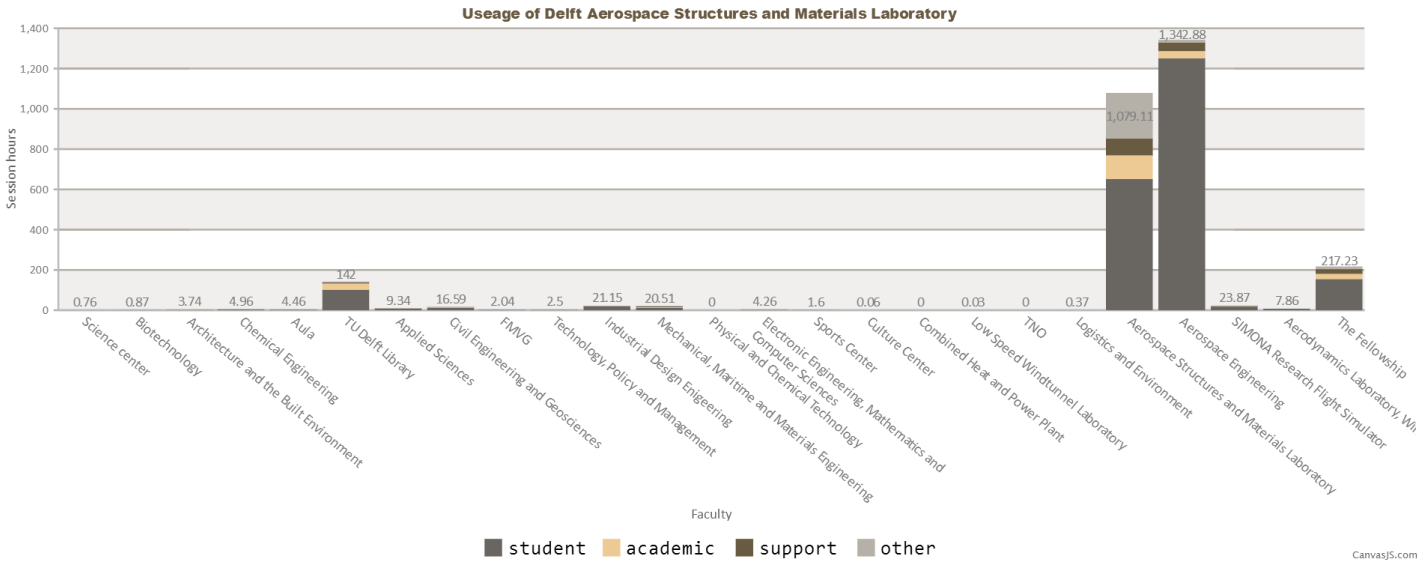


Figure A.63: Information of user group of building 61

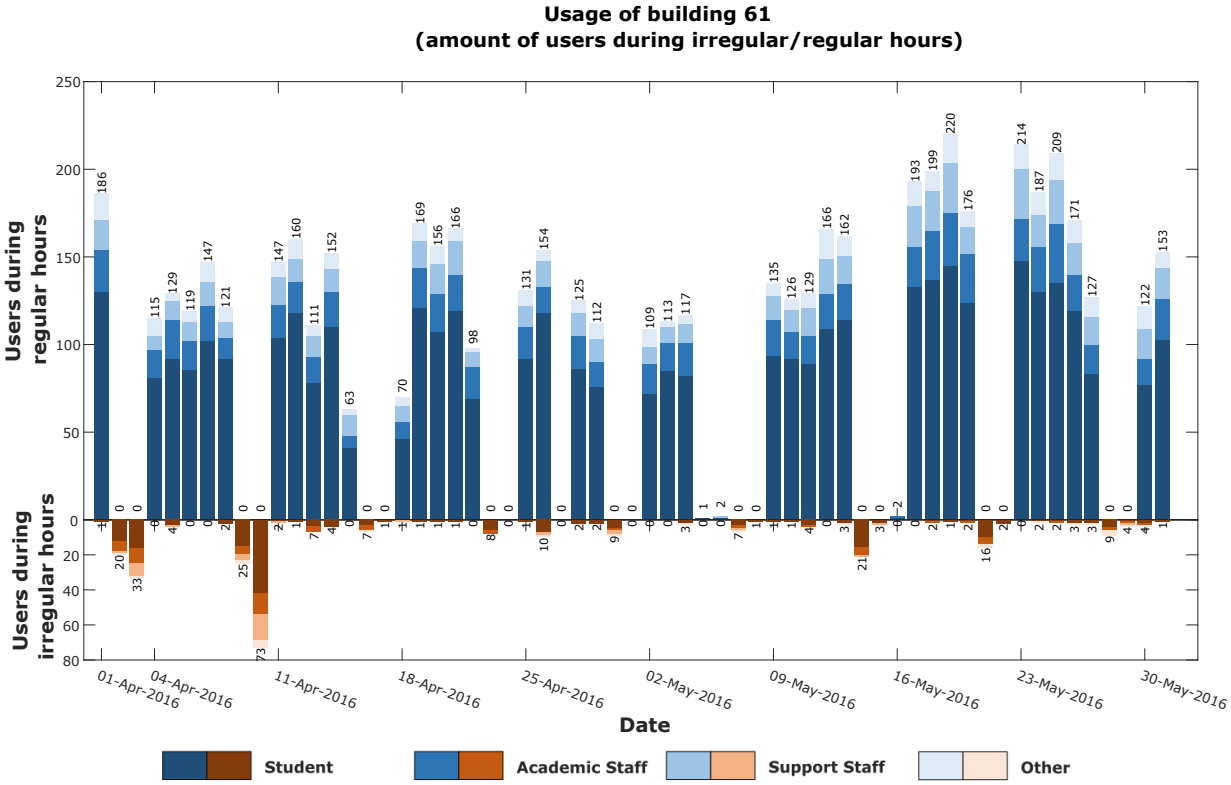


Figure A.64: Usage of building 61, amount of users present, divided by type of user

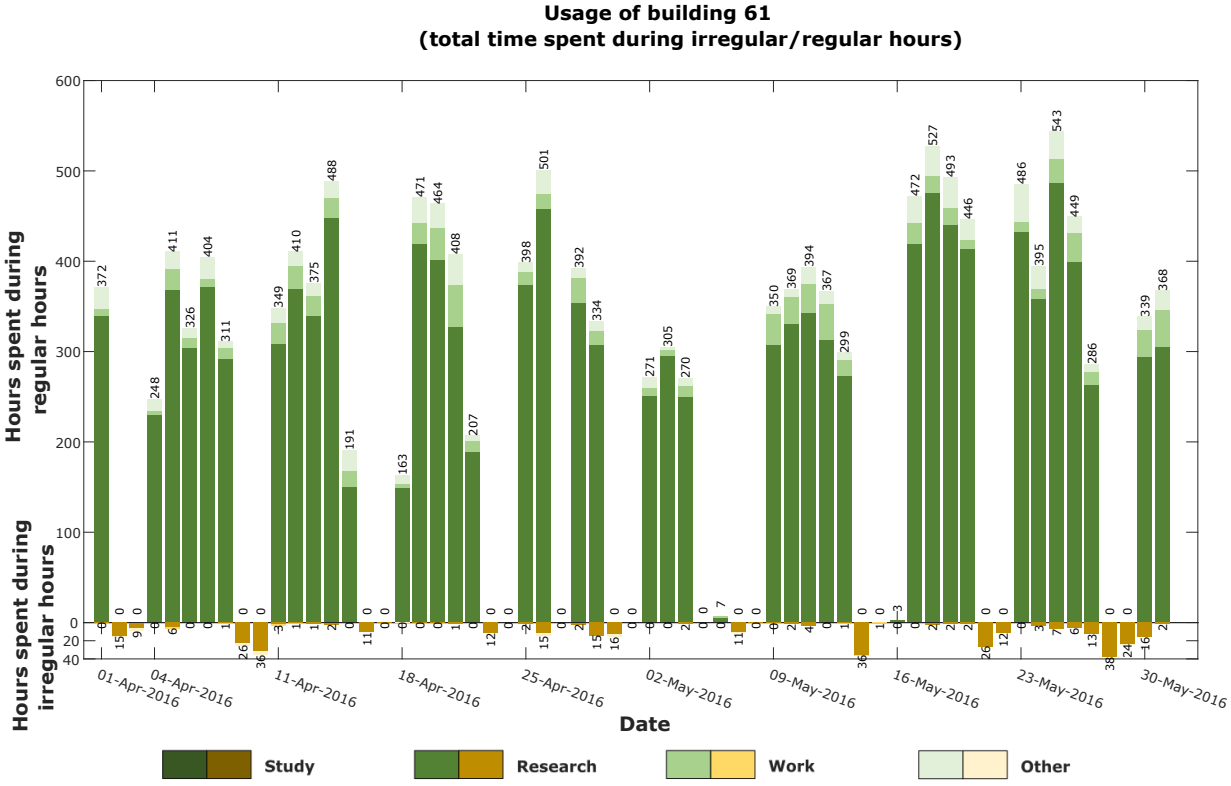


Figure A.65: Usage of building 61, the total time spent by users divided into activities

A.23. Building 62: Faculty of Aerospace Engineering

Type of building: faculty building
Opening hours: Monday to Friday 07:00 - 22:00 and Saturday 10:00 - 16:00
Exceptions:none

Users and time spent during irregular/regular hours
There are about 50 users found on every Saturday in the building and each user spends around 2.5 hours to 3 hours; most of them are studying.
On average, the daily user number of Aerospace Engineering is 1000 to 1200 after the exam and is 800 during the exam in regular hours. The time they spend during regular hour is three to four hours per day. Usually, user number on Fridays is slight lower (150 lower) than on other weekdays.
About 50 people use the building in irregular hours per day after the exam; on average, each of them spends 1.5 to 2 hours. Less people use the building in irregular hours during the exam week; the average time they spend per day is also 1.5 to 2 hours per day.
An increment of user number in irregular hours is found between 17th May to 21th May; on Wednesday, Thursday, Friday and Saturday of that week, there were 92, 69, 167 and 429 users found respectively using the building for about one hour. The proportion of academic staff working on 21th May is larger than other days.
The daily user number of Aerospace Engineering is 1000 to 1200 after the exam and is 800 during the exam in regular hours. On average there will be 50 users studying in the building for 2.5 hours on Saturday. About 50 people use the building in irregular hours in weekdays.

User group information
The Faculty of Aerospace Engineering is mainly used by people from itself (in total 15,312 hours, on average 255 hours per days by all the users); therein, student (63.5%), academic staff (18.7%), supports staff (11.31%) and others (6.46%). The building is barely used by people from other buildings. People from the Fellowship and TU Delft Library use the building more often because they need to do study at the library and use better computer for computation.

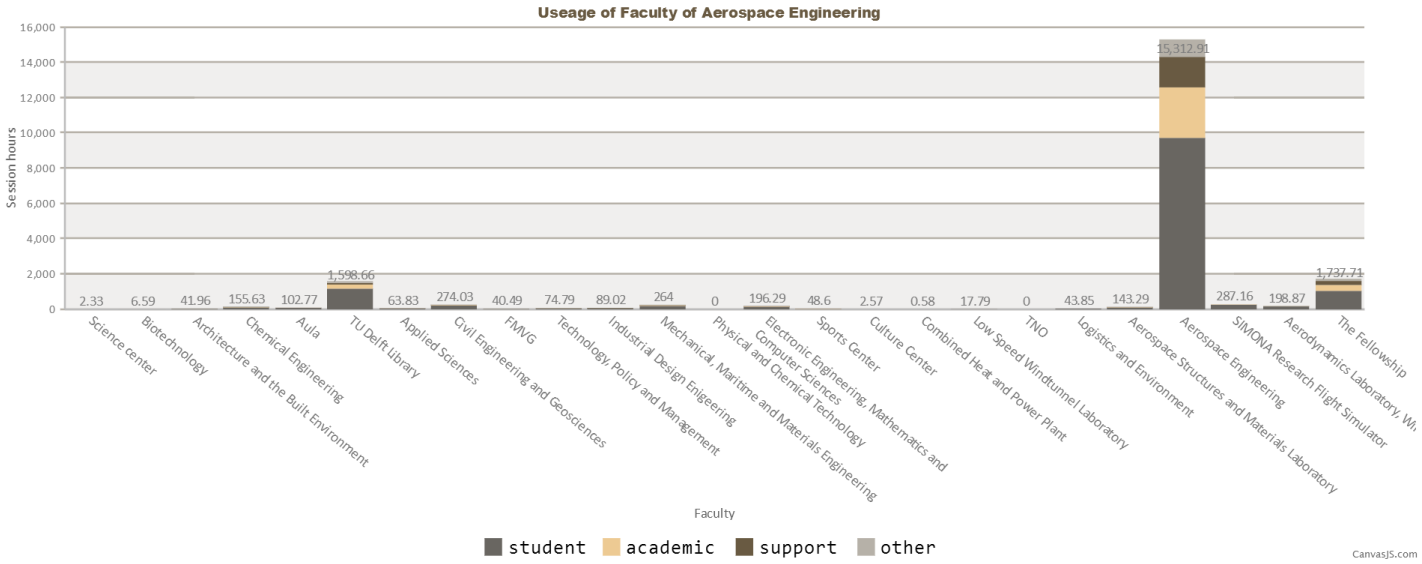


Figure A.66: Information of user group of building 62

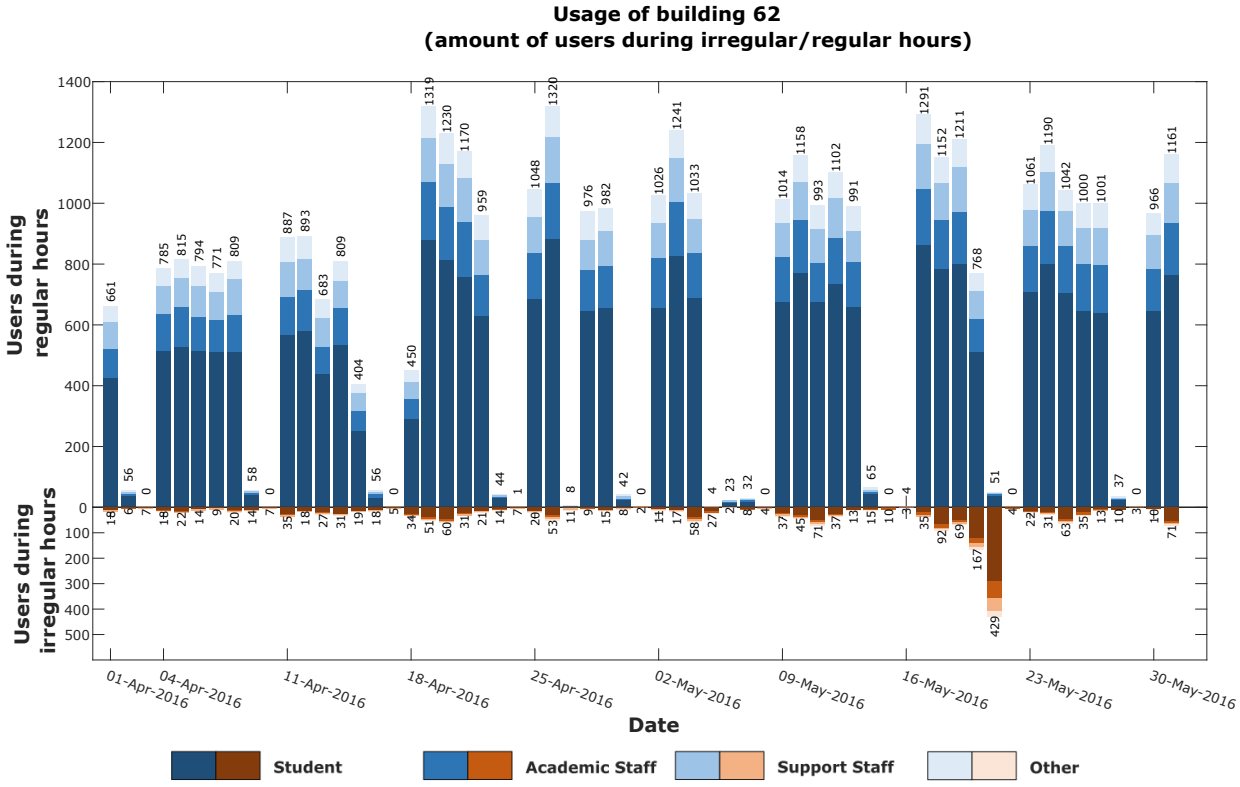


Figure A.67: Usage of building 62, amount of users present, divided by type of user

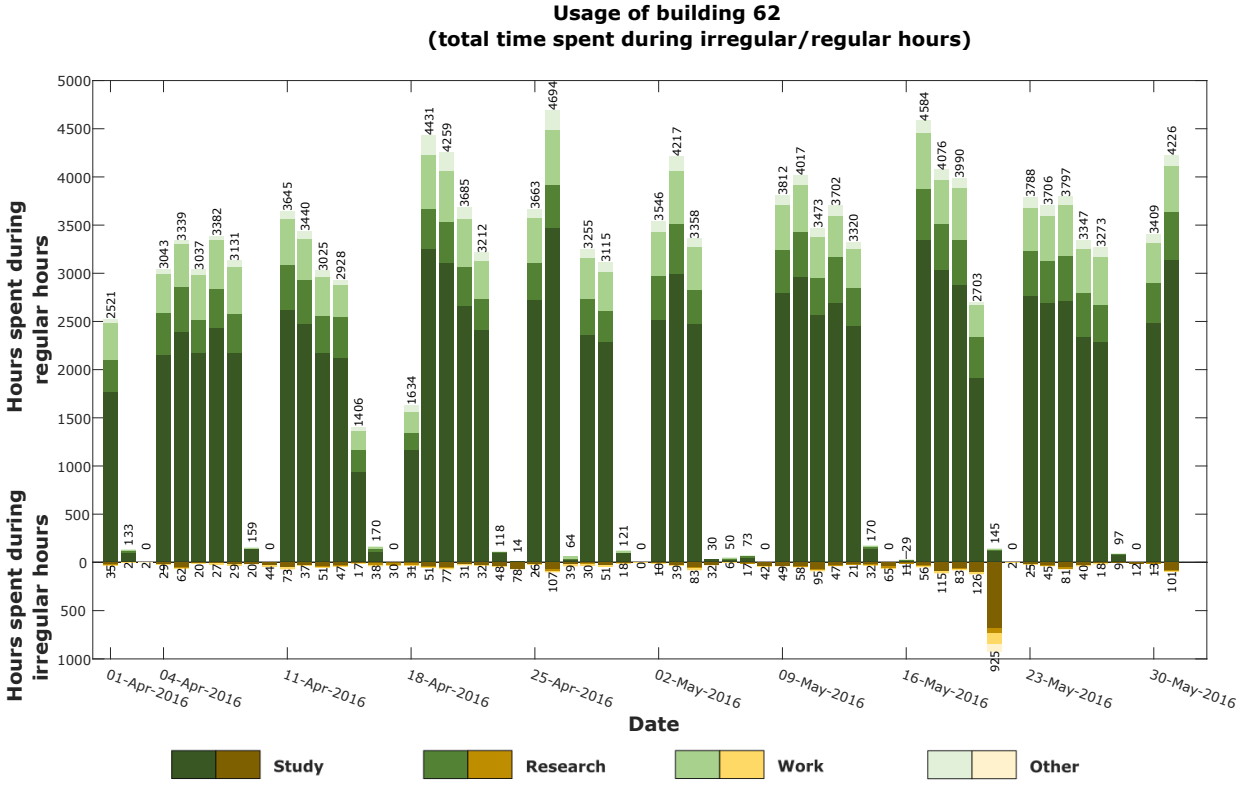


Figure A.68: Usage of building 62, the total time spent by users divided into activities

A.24. Building 63: SIMONA Research Flight Simulator

Type of building: research building
Opening hours: Monday to Friday 08:00 - 18:00
Exceptions: authorised campuscard after closing

User number and time spent during irregular/regular hours
The daily user number during exam week is about 60 during regular hours and it varies from 6 to 31 during irregular hours but is around 25 in general in weekdays.
In weekends, during exam week, around 10 users use the flight simulator laboratory; the session time varies from half an hour to three hours. The user number increased slightly from 70 to around 90 after the exam; the peak was found on 3rd, 4th and 13th of May which is 114 users. The session time varies but is around 2.5 hours to 3 hours per person per day during regular hours in weekdays. In general, users on Saturdays are more than users on Sundays. A drop of user number also can be found on every Friday, on 15th of April (Friday), 10th of May (Friday) and 27th of May (Friday), the user number was only 50%, 50% and 70% respectively of the average daily value.
Most of the users are students; compared with the other buildings, the proportion of academic staff is higher. However, studying still takes the largest part of all the activities in the building.
SIMONA Research Flight Simulator laboratory has 220-270 daily users during regular hours and 25 daily users during irregular hours in weekdays. On Saturdays, usually there are 16-20 people working for 1.5 to 2 hours; on Sundays, usually there are less than 4 people there and work for one to three hours.

User group information
The building of SIMONA Research Flight Simulator is mainly used by people from itself (in total 1,050 hours, on average 17.5 hours per days by all the users); therein, student (75.7%), academic staff (19.22%), supports staff (1.8%) and others (3.28%). The building is also used by people from Aerospace Engineering and the Fellowship which makes sense because the lab is related to aerospace. Besides, the laboratory is barely used by people from other buildings.

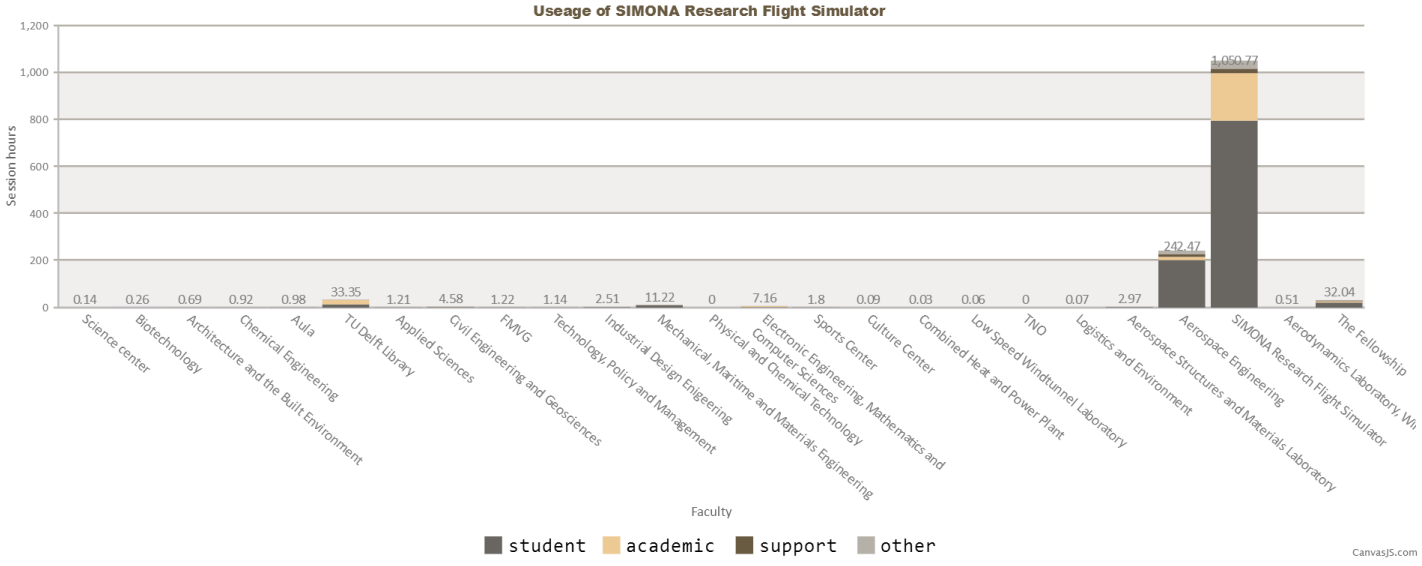


Figure A.69: Information of user group of building 63

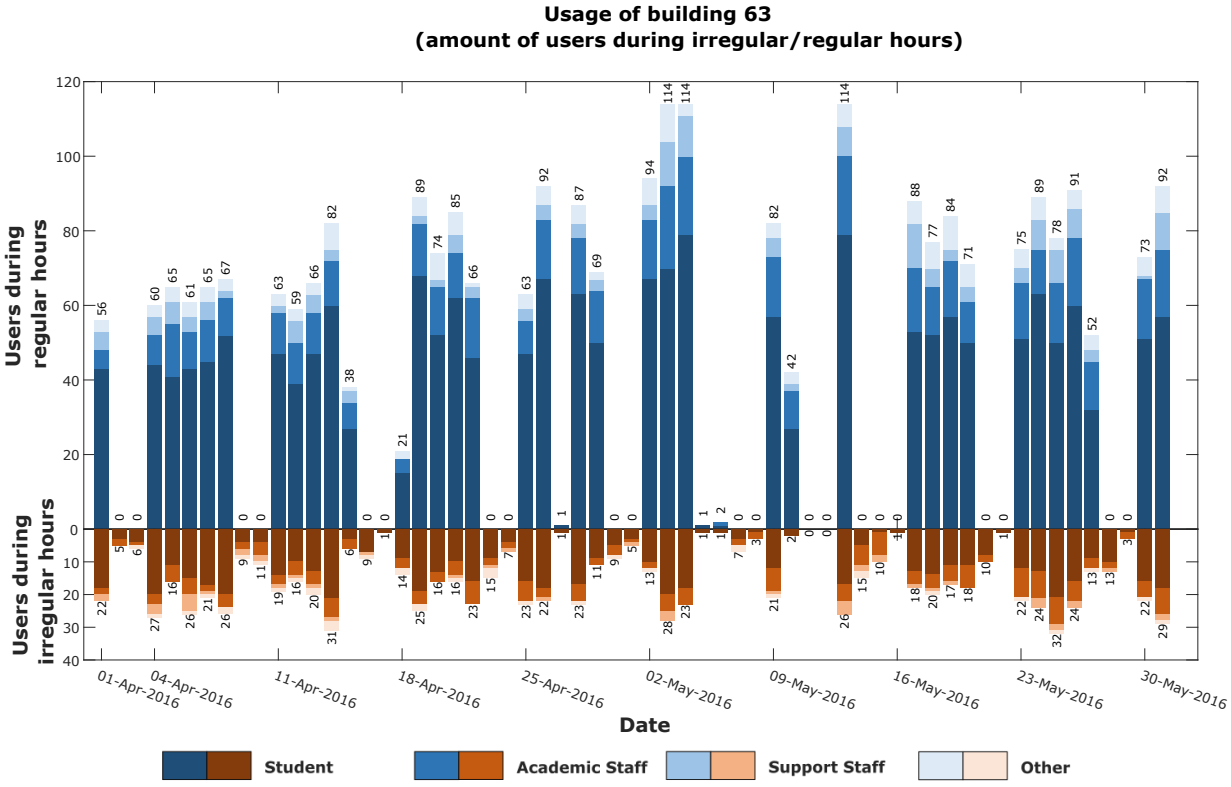


Figure A.70: Usage of building 63, amount of users present, divided by type of user

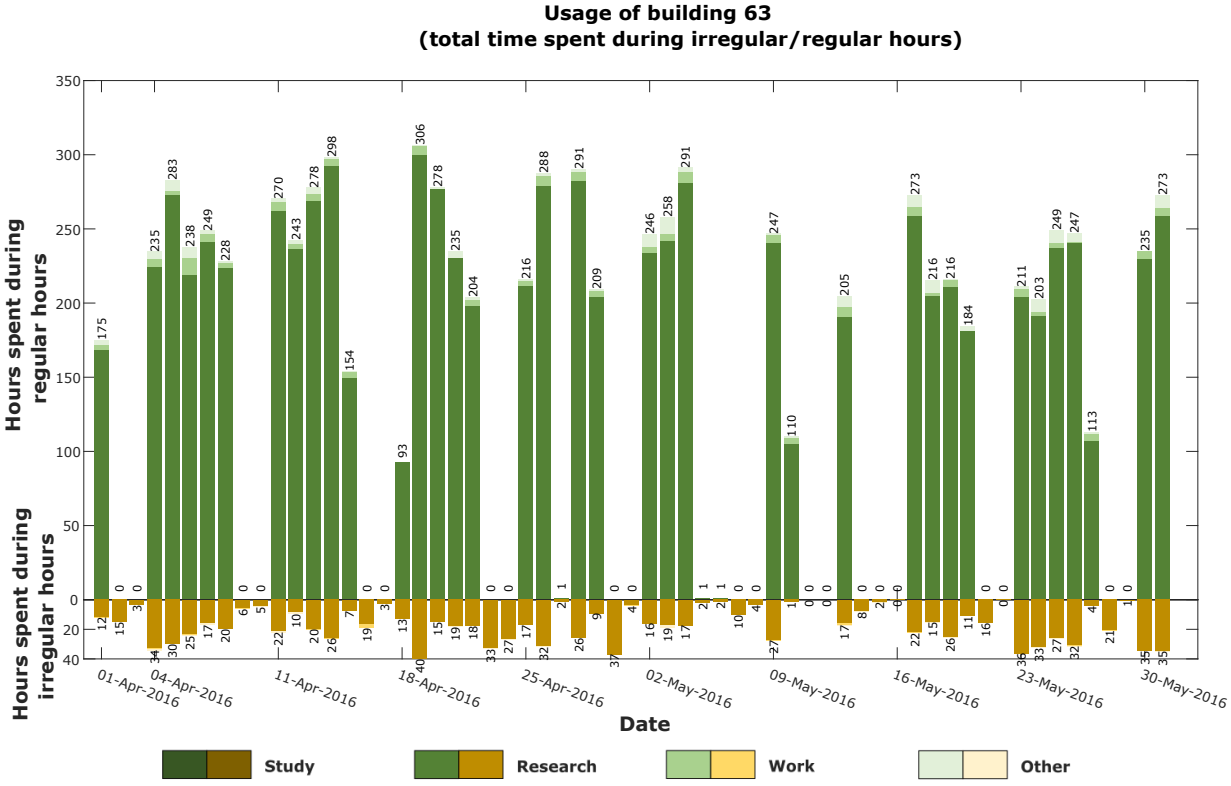


Figure A.71: Usage of building 63, the total time spent by users divided into activities

A.25. Building 64: Aerodynamics Laboratory, Windtunnels

Type of building: research building
Opening hours: Monday to Friday 07:00 - 22:00 and Sunday 10:00 - 16:00
Exceptions: authorised campuscard after closing

User number and time spent during irregular/regular hours
In general, less than 5 people work in irregular hours for around or less than one hour in weekdays. About 10 people can be found on each Sunday work for a relatively long period (about 2 to 4 hours), it is due to the Aerodynamics Laboratory and wind tunnel is open on Sunday for 8 hours. However, even the building does not open on Saturdays, there are still more users found on Saturdays than on Sundays and their session time is around 2 hours. The daily user number in weekdays during regular hours before 9 May is around 85 to 90, except 15th April (the Friday of the exam week, 36 users) and 18th April (the first day of the new quarter, 31 users); however, the user number increased sharply to 102 on the second day of the fourth quarter possibly due to the new courses started on that day. The peak of the user number in regular hours can be found on 10th May which is 153 users and the peak of the user number in irregular hours can be found on 2nd April which is 20 users. In general, users in this laboratory increased since 9th May. The proportion of academic staff and undefined people are relatively high compared with the other main faculties probably due to academic staff doing experiments in the laboratory. According to the dashboard, the laboratory is mostly used by students and staff from Aerospace Engineering.

User group information
The Aerodynamics Laboratory and Wind-tunnels is mainly used by people from itself (in total 2091 hours, on average 34.85 hours per days by all the users); therein, student (61%), academic staff (22.31%), supports staff (7.02%) and others (9.6%). The building is also used by people from Aerospace Engineering which makes sense because the lab is related to aerospace. Besides, the laboratory is barely used by people from other buildings.

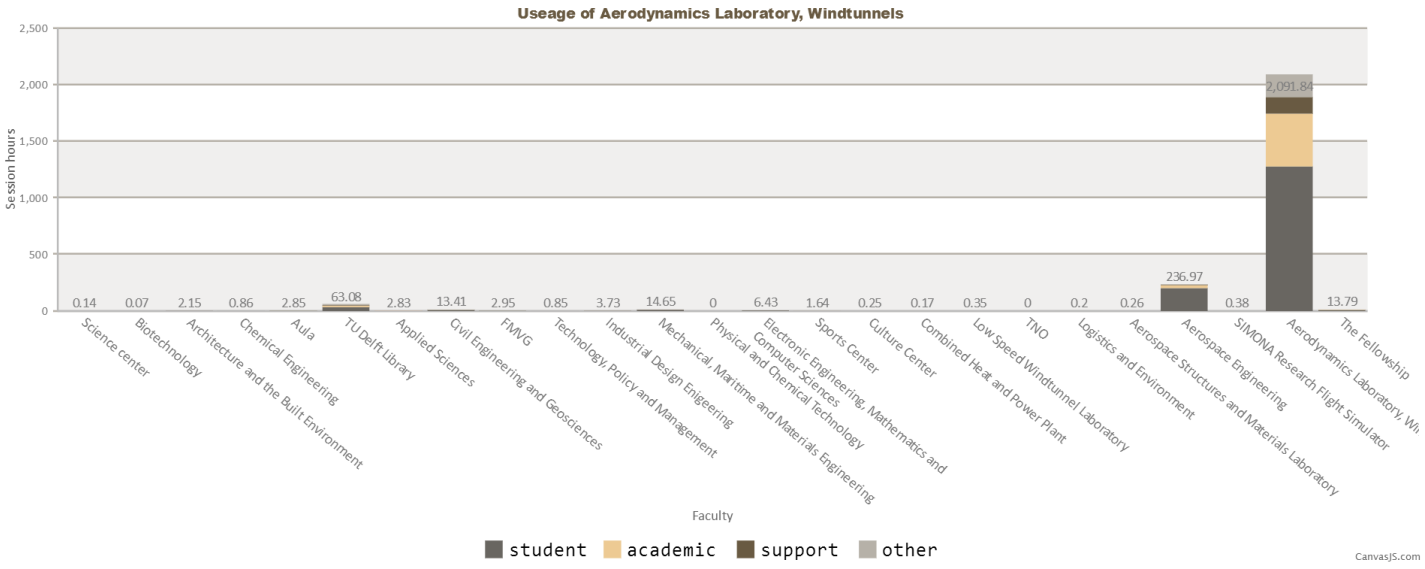


Figure A.72: Information of user group of building 64

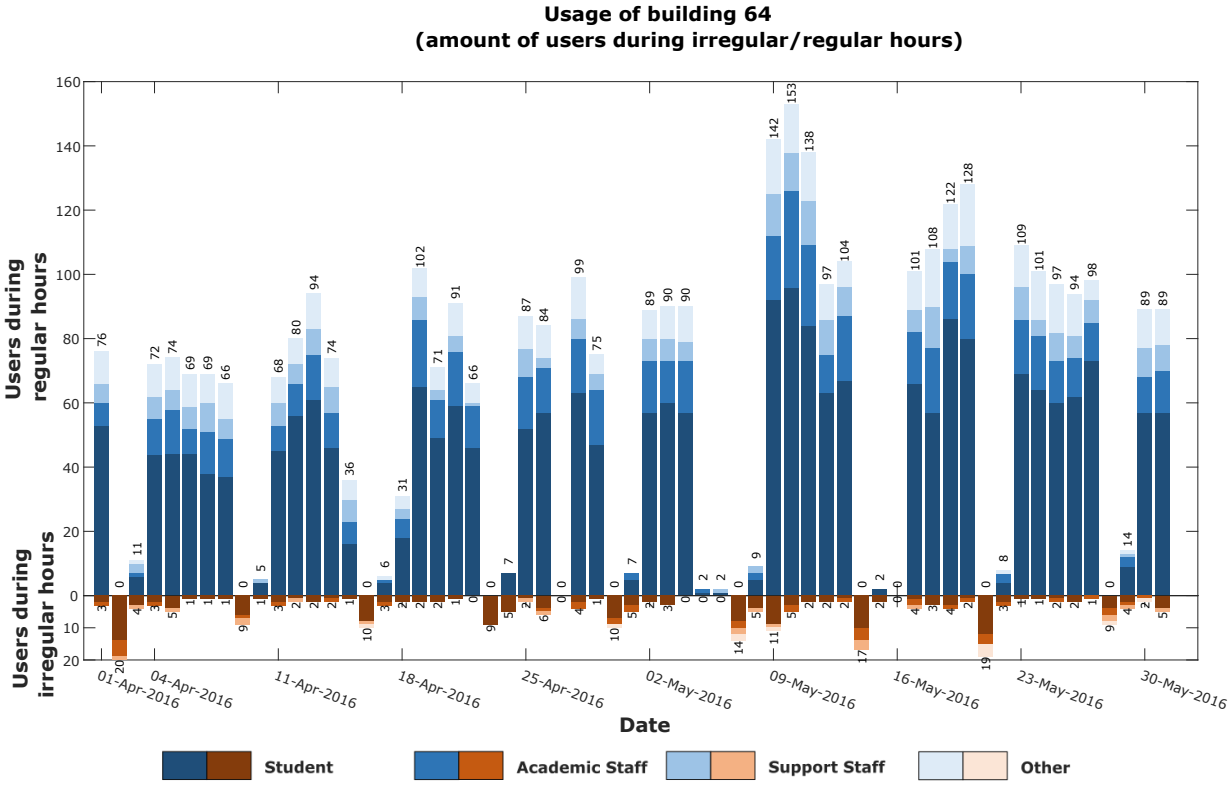


Figure A.73: Usage of building 64, amount of users present, divided by type of user

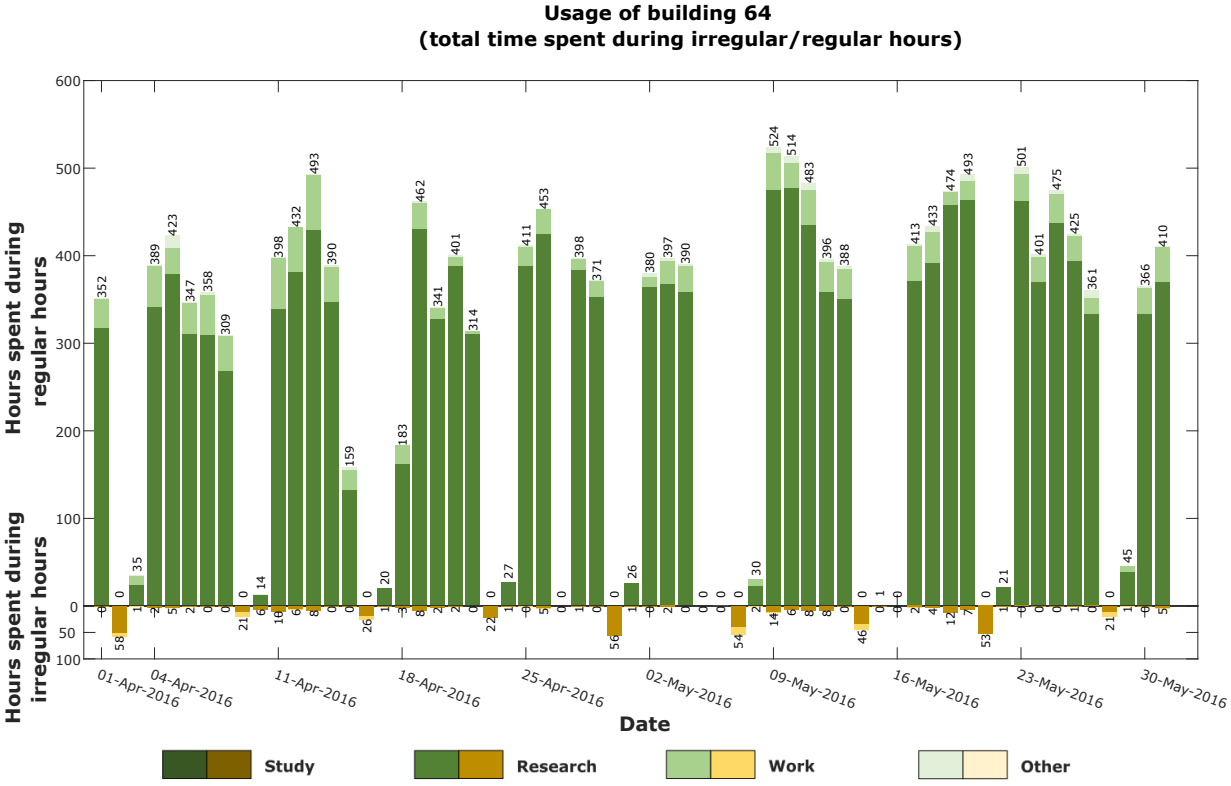


Figure A.74: Usage of building 64, the total time spent by users divided into activities

A.26. Building 66: The Fellowship

Type of building: educational building
Opening hours: Monday to Friday 08:00 - 20:00
Exceptions: none

User number and time spent during irregular/regular hours
It can be easily found that a larger amount use of the Fellowship in irregular hours was during the exam week (3351 users); after the exam week, there were only 215 users (6.4% of what during the exam week) found after the exam. The daily user number during the exam in regular hours is 370 to 400 and dropped sharply on the Friday of exam week due to the end of all the exams. After the exam, the number increased about 50% in regular hours but the user number dropped sharply in both weekdays and weekends as well. The peak of user number in regular and irregular hours can be found on 21th April which was 1077 and on 10th April which was 593. In weekends during the exam week, the average time one user spent in irregular hours is about 6.5 hours; during the weekdays, the average time per person is about 2.5 hours. The time users spend In regular hours in weekdays after exam varies from 3 to 4 hours per day. The time user spent in irregular hour in weekdays after exam week also varies but mostly shorter than half an hour. The daily user number in regular hours in weekdays during the exam week is around 370 to 400 while what in irregular hours is around 140 (weekdays) and around 500 (weekends); the number after exam week in regular hours increases to around 700 to 800 while what in irregular hours drops sharply to around 10.

User group information
The Fellowship is mainly used by people from itself (in total 11,387 hours, on average 189.8 hours per days by all the users); therein, student (64.84%), academic staff (14.27%), supports staff (14.18%) and others (6.71%). The building is also often used by people from TU Delft Library, Aerospace Engineering, Mechanical, Maritime and Material Engineering and Civil Engineering and Geoscience which makes sense because the fellowship provides these faculties with better computers and facilities.

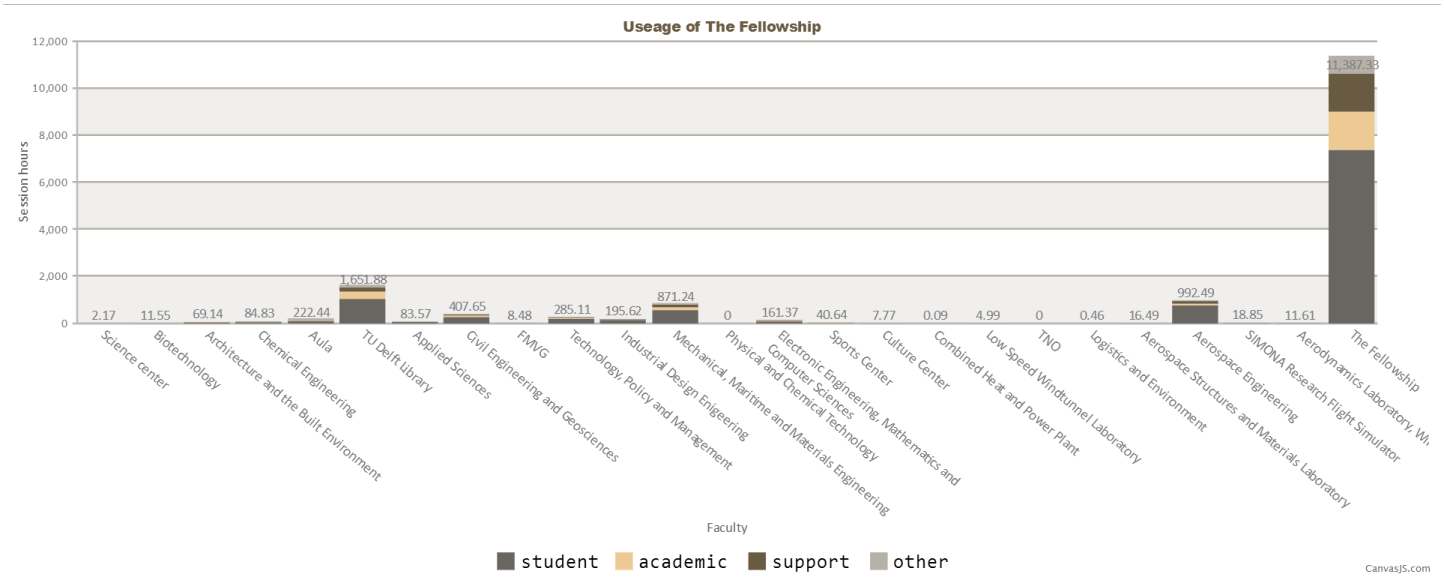


Figure A.75: Information of user group of building 66

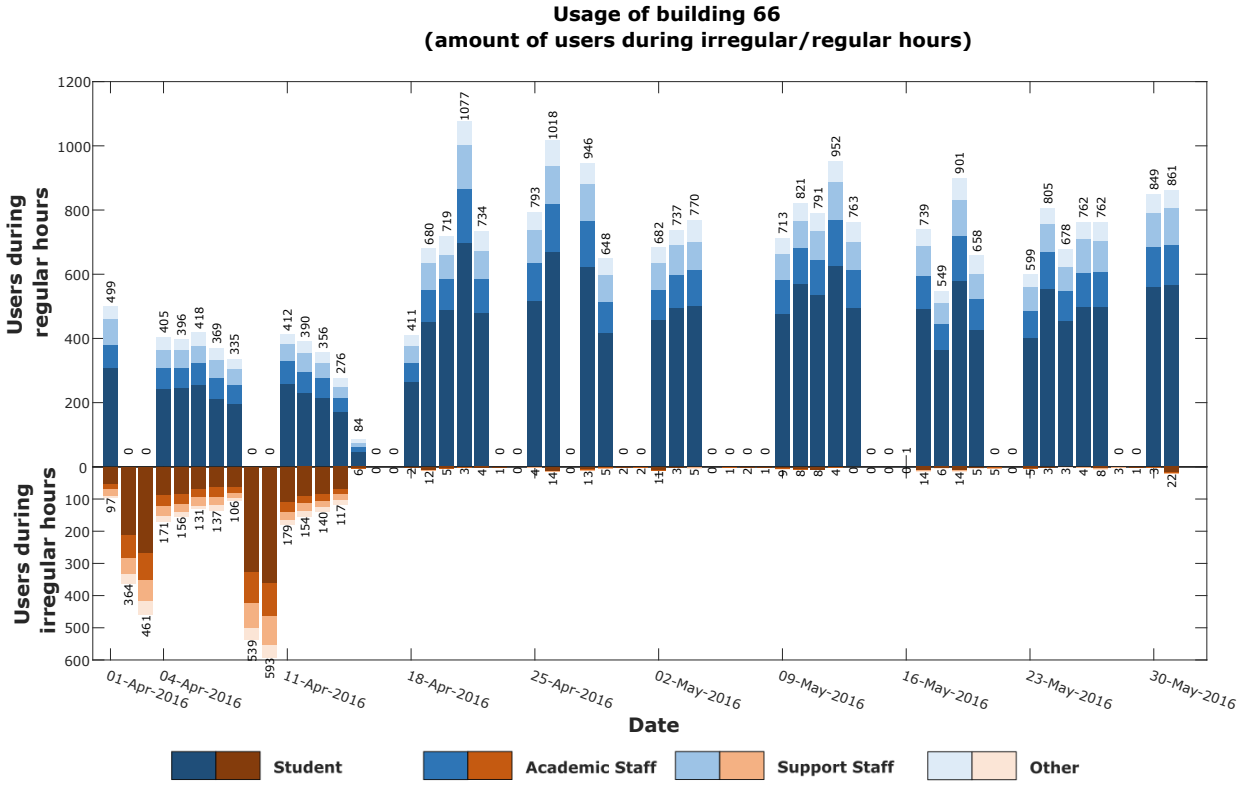


Figure A.76: Usage of building 66, amount of users present, divided by type of user

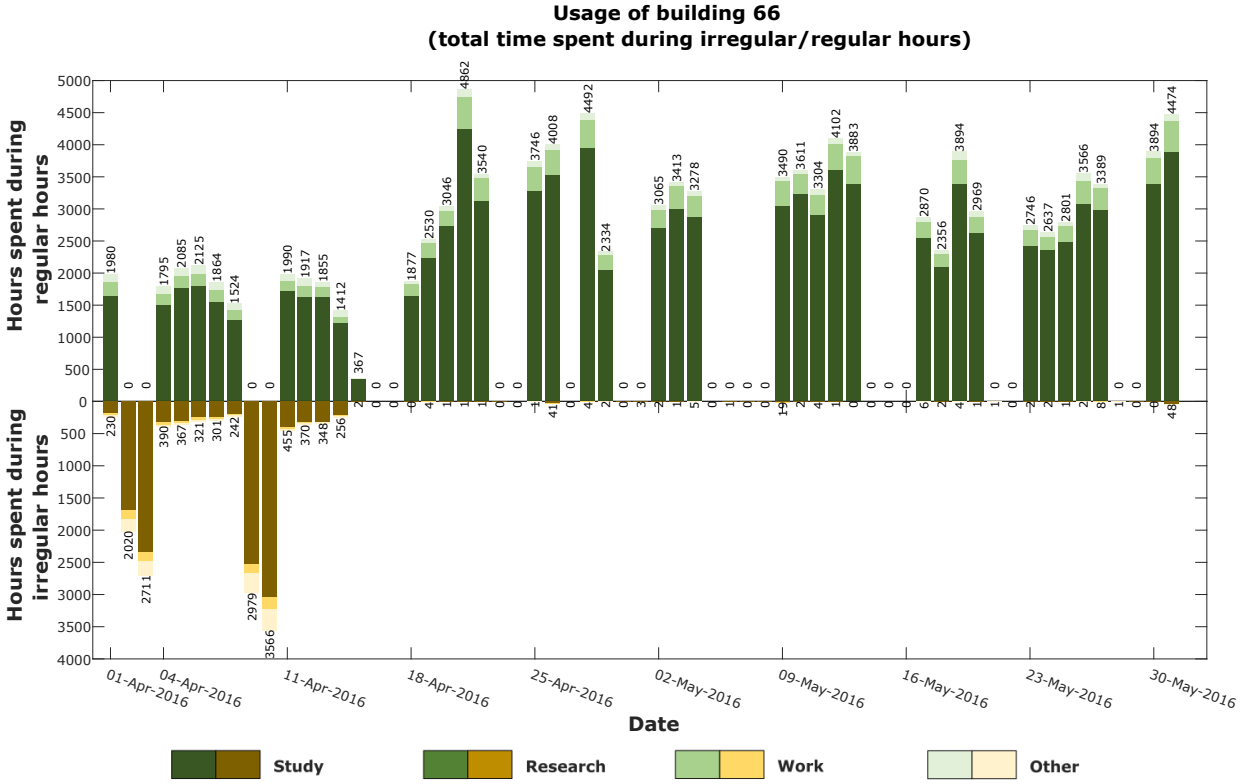


Figure A.77: Usage of building 66, the total time spent by users divided into activities

B

Questionnaire

Geomatics Synthesis Project Questionnaire

During the final quarter of the first year of the master program of Geomatics for the Built Environment of the Delft University of Technology (TU Delft), groups of students work intensively on the geomatics synthesis project (GSP). The aim of this project is to provide the TU Delft with information about the use of the facilities on the campus during the irregular hours. The irregular hours are specified as the hours outside of the opening hours of the buildings.

The questionnaire results will be utilised as training samples and for the validation of the results of our project.

Your participation will be appreciated.

* Required

1. **1) What is your occupation?** *
   Mark only one oval.
   - ◯ Student
   - ◯ Academic Staff
   - ◯ Support Staff
   - ◯ Other

2. **2) To which faculty do you belong? In case you belong to more than one faculty, select the one you spend the most time at.** *
   Mark only one oval.
   - ◯ Aerospace Engineering (AE)
   - ◯ Applied Sciences (AS)
   - ◯ Applied Sciences (AS), Environmental Biotechnology (EBT)
   - ◯ Applied Sciences (AS), Chemical Engineering (ChemE)
   - ◯ Architecture and the Built Environment (Arch)
   - ◯ Civil Engineering and Geosciences (CEG)
   - ◯ Electrical Engineering, Mathematics and Computer Science (EEMCS)
   - ◯ Industrial Design Engineering (IDE)
   - ◯ Mechanical, Maritime and Materials Engineering (3mE)
   - ◯ Technology, Policy and Management (TPM)
   - ◯ Other: .....

3. **3.a) Do you use University facilities after closing hours? (e.g. by acquiring special permission or using card access)** *
   Mark only one oval.
   - ◯ Yes, less than one hour per week.
   - ◯ Yes, between one and seven hours per week.
   - ◯ Yes, more than seven hours per week.
   - ◯ No

4. **3.b) In case you selected yes, what activities do you perform after the closing hours?** *
   Check all that apply.

   - [ ] Study
   - [ ] Work (non-research work)
   - [ ] Work (research or lab activities)
   - [ ] Recreational activities
   - [ ] Other: .....

5. **4.a) Do you use the library during the week?**
   Mark only one oval.

   - ◯ Yes, less than four hours per week.
   - ◯ Yes, between four and sixteen hours per week.
   - ◯ Yes, more than sixteen hours per week.
   - ◯ No

6. **4.b) Do you use the library on weekends?**
   Mark only one oval.

   - ◯ Yes, less than three hours.
   - ◯ Yes, between three and ten hours.
   - ◯ Yes, more than ten hours.
   - ◯ No

7. **5.a) Do you use the sports and culture centre during the week?**
   Mark only one oval.

   - ◯ Yes, less than three hours per week.
   - ◯ Yes, between three and ten hours per week.
   - ◯ Yes, more than ten hours per week.
   - ◯ No

8. **5.b) Do you use the sports and culture centre on weekends?**
   Mark only one oval.

   - ◯ Yes, less than three hours.
   - ◯ Yes, between three and ten hours.
   - ◯ Yes, more than ten hours.
   - ◯ No

9. **6) Do you use other university facilities on weekends?**
   Mark only one oval.

   - ◯ Yes, less than three hours.
   - ◯ Yes, between three and ten hours.
   - ◯ Yes, more than ten hours.
   - ◯ No

10. **7) Where do you usually have your lunch?**
Mark only one oval.

- ◯ Aula
- ◯ Faculty
- ◯ Other

11. **9) How many University buildings do you usually visit per week?**
Mark only one oval.

1	2	3	4	5	6	7	8	9	10
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

C

Queries manual

Pre-requisites:

PostgreSQL must be installed

Remote database of wifilog access with writing privileges
OR
Localhost database to store the tables with access to remote database of wifilog with reading privileges

PostgreSQL steps

First step:
Either run directly on remote database
1) Skip query 1. DBLINK_server2local_wifilog

Or run on local database
1) Run 1. DBLINK_server2local_wifilog
   →wifilog is copied from remote to local

Next step:
2) Run 2. Processing
   →wt3_building is created
   →wt3_wifilog is created
   →wt3_apname is created
   →wt3_maploc is created
   →wt3_bno is created
   →wt3_campus is created
3) Run 3. Profiling
   →wt3_userstatistics is created
4) Run 4. Occupation
   →wt3_trainingsets is created
   →wt3_users is created
5) Run 5. Activity
   →wt3_bno_activity is created

D

Dashboard manual

Preparation

1. Create a work space in your geoserver, publish the “buildings” shapefile in your geoserver and name the layer “buildings” (see the following link about how to publish a shapefile). Publishing a shapefile in geoserver:
http://docs.geoserver.org/latest/en/user/gettingstarted/shapefile-quickstart/index.html
2. PHP is used in the application; therefore, to run the application, the server should be installed with PHP.
3. The following steps have to be done after reading the database manual.
4. The function of “Weekly building occupancy” is based on table “reg_bno” and “irre_bno” which can be obtained by running “ir_re_bno.sql”.
5. The function of “hourly user count” is based on table named “hourly” in your local database. The query for creating the table is “hourly.sql” and is packed in the file.
6. The function of “Building usage” is based on table “b_use” which can be obtained by execute “building_use.sql”.
7. Open the application folder in Microsoft WebMatrix or other web development tool with PHP installed (Web Matrix can be downloaded from: https://www.microsoft.com/web/webmatrix/; a YouTube tutorial of how to install PHP starter: https://www.youtube.com/watch?v=isZl1HzTaK4).
8. Open “reg.php”, “hourly.php”, “build_use.php”, search and replace string ”host=localhost dbname=synthesis port=5432 user=postgres password=###” to your host, database name, port, username and password.
9. Open “main.js” under “src” folder, search and replace all “GEOWEB1” to you the name of the work space you created in geoserver.
10. If you are using Chrome, go to
http://tinyurl.com/kd7zav3 and install the plugin called “Allow-Control-Allow-Origin” and enable it as shown in Figure D.1.

Figure D.1

11. If you are using Firefox, go to https://addons.mozilla.org/en-US/firefox/addon/cors-everywhere/ download and install “CORS Everywhere” and enable it as shown in Figure D.2 (These two plugins allow your localhost to exchange data with your geoserver).

12. Launch the index.html in preferred browser.

Figure D.2

User guide

1. The user-interface should look like what is shown in Figure D.3.

Figure D.3: User interface

2. Open the sidebar by either click the building button or click on the campus building colored with green.
3. The sidebar looks as shown in Figure D.4. There are three functions developed, weekly building occupancy, hourly user statistics in a specified day at either building level or map location level (comparison is included) and user occupation of a selected building in a specified week

Figure D.4: Sidebar

4. Function 1: Hourly user statistics in a specified day at either building level or map location level

The function counts user number per hour in a specified day at either map location or building level. The change of occupancy of the building or map location along with the time series can be easily obtained and clearly visualized. Activities and events can be presumed according to the change of occupancy or the comparison of occupancy.
- Select a building that you are interested in from the drop down box.
- Select “no” from map location drop down list if you do not want to specify a map location and the result will be at a building level.
- If you want the result at map location level, specify a map location.
- Specify one day in one week.
- Repeat the above steps in the comparison section for another map location or another day.
- Click the button: ”Compare usercount”
An example shows the ground floor of Faculty of Architecture and Built Environment on the Friday and Saturday of week 16 is shown in Figure D.5 (same location different time).

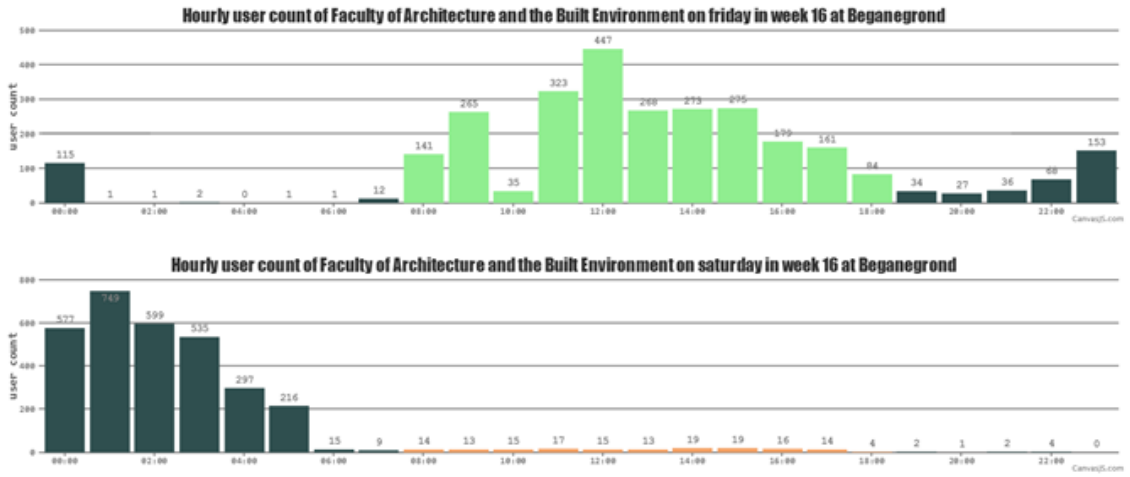


Figure D.5: User count on the ground floor of BK on the Friday of week 16

5. Function 2: User occupation of a selected building in a specified week
This function shows the main users of a building in addition with what are the users’ main building and what are the users.
- Select the interested building by clicking the campus building map or select from the drop down list.
- Specify a week.
- Click ”User occupation” to show results.
An example of sports center is shown in Figure D.6.

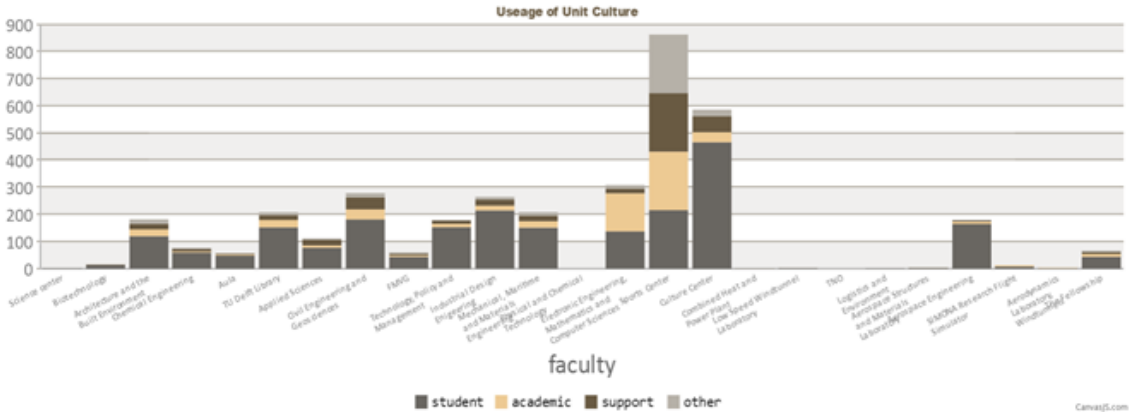


Figure D.6: Occupation of the users go to the sports center

E

DID-B1 Project Plan

IJ.D.G. Groeneveld
R. Sulzer
E. Theocharous
M.S. Tryfona
O.T. Willems
Y. Xu

Project Plan

Synthesis Project 2016 – SP Big Brother

Project Plan

Synthesis Project – SP Big Brother

By

IJ.D.G. Groeneveld
R. Sulzer
E. Theocharous
M.S. Tryfona
O.T. Willems
Y. Xu

in partial fulfilment of the requirements for the Synthesis Project of

Master of Science
in Geomatics for the Built Environment

at the Delft University of Technology,
to be presented on Monday April 25, 2016 at 09:00 AM.

Project Coordinator:	Stefan van der Spek
Team Coach:	Wilko Quak

An electronic version of this report is available at Blackboard

Contents

A. Executive Summary

The aim of the TRACK-id project is to provide the Facility Management and Real Estate (Facilitair Management & Vastgoed (FMVG)) of the Technical University of Delft (TUD) with information about the use of the facilities on the campus during irregular hours. The irregular hours are specified as the hours outside of regular office hours and education hours (lectures and laboratory work), as well as the hours during weekends. With this information and possibly resulting change of the usage of facilities, the people that move around the campus, such as students, researchers and other staff can benefit of a better campus and more centralized facilities. The information that we will derive is of key importance to FMVG, in order to provide some recommendations and policy support concerning the use of real estate. With this they can provide more efficient and centralized use of this campus' real estate during irregular hours.

In order to get this valuable information, FMVG will provide the Wi-Fi data from all the access-points that are installed in and around the buildings on the campus of the TU Delft. This data, together with reference data, such as the map of the campus, and usage of spaces will be used during processing. Depending on the needs during the different stages of the process, the available support staff and necessary technology will be accessed and relied upon. During the project we will divide the Wi-Fi tracking dataset into two parts: the dataset during regular hours and the dataset during irregular hours. The project will delve into the different information that can be obtained and derived from the dataset. More specifically, statistical analysis will be applied for a) the identification of users according to a set of profiles, including, but not limited to, students, employees and visitors, b) the different devices and the corresponding use that a user(group) has, and c) the behaviour of the different user (are they attending lectures, lunching, studying in library, in laboratory, etc.). An extensive analysis will result in an insight in the behaviour of people. The aforementioned analysis will be on both the subset of regular and the irregular hours, in order to compare the two subsets. An important part of the project is the validation of all the assumption that are made to get the results. The results will be validated a small part of the dataset that includes the users operating within the Faculty of Architecture, in order to determine the quality of the outcome. Finally, the (validated) results will be visualised using appropriate software and the end-product will be insight in the behaviour of users and a set of recommendations, according to the needs of the clients. During the process, the requirements of FMVG, the legal restrictions, and the potential competitors will be taken into consideration.

The baseline review includes the project plan, the requirements specification with an inclusive requirement discovery tree and a rich picture that provides a visualisation of the position of the project in the context of topics, stakeholders and threats versus opportunities. The project plan contains the Organizational Breakdown Structure (OBS), the Work Breakdown Structure (WBS), the Work Package Descriptions (WPD), the Project Logic Diagram, and the schedule of the project activities, showing the project phasing and the planning of deliverables (GANNT chart).

B. Project

1.0. Introduction

This document is the project plan for the team “TRACK-id” of the 2016 Synthesis Project (SP) as part of the master program Geomatics. In the GEO1101 SP we combine and apply all the previously acquired knowledge in the field of Geomatics to work on a real-word group research project. This year’s topic of the project is monitoring occupation, flows, and behaviour patterns using Wi-Fi tracking. TRACK-id focuses on the activities, i.e. behaviour, of users on the TU Delft Campus during regular and irregular hours. The objectives of this document is to define the process and organisation of our research.

In the following chapters the organizational (Section 1.1) and work breakdown structure (Section 1.2) are described, together with a work package description (Section 1.3). In Section 1.4 the flow of activities is shown in a project logic diagram, and the schedule of the project is provided in a GANNT chart in Section 1.5.

1.1. Organizational Breakdown Structure

In order to work productive, each group member is assigned organizational responsibilities. The organizational tasks are assigned with the help of the personal A4 and preferences of every group member.

The Organizational Breakdown Structure (OBS) (see Figure 1) shows the communication lines for this project. The project team communicates with the client (FMVG), the team coach (Wilko Quak), and the project coordinator (Stefan van der Spek). Currently there is no communication with external partners/actor, however it is expected that this will be the case during the progression of the project. In the project team different responsibilities/tasks are identified. Each team member will be responsible for at least one task based on the personal A4. As the team consists only of six people most task will be done together with part of or the entire team.

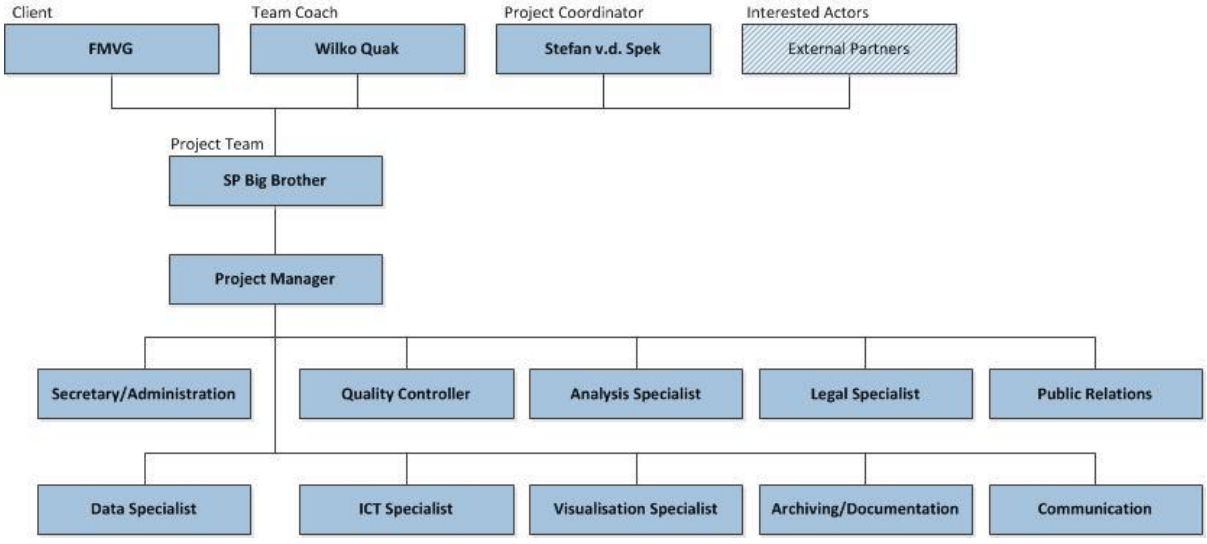


Figure 1: Organizational Breakdown Structure

1.2. Work Breakdown Structure

Before the project goes into the technical part and work mode, all task, that have to be done to reach the final product need, are identified .

Figure 2 shows the work breakdown structure of the project.

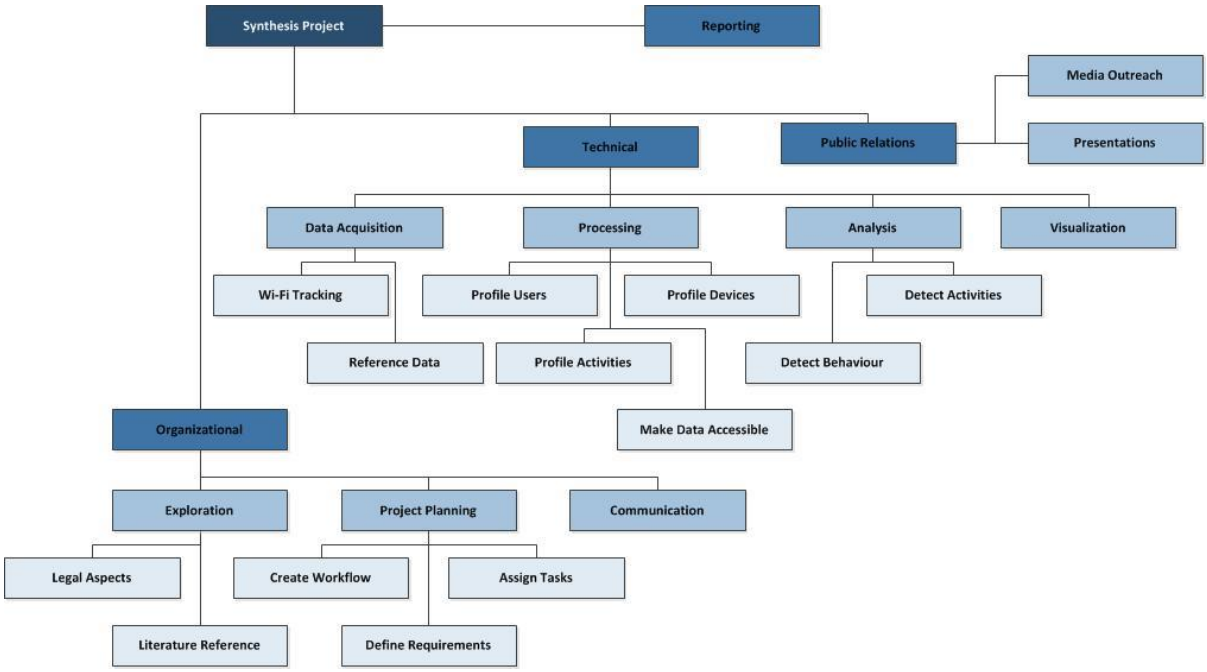


Figure 2: Work Breakdown Structure

1.3. Work Package Description

Organizational: This work package focuses on the organizational part of the project. The overall scope of the project is defined in this phase. This includes exploration of relevant literature and legal aspects (opportunities and limitations) of the project. During project planning the requirements, mainly from the stakeholders, are identified. Additionally, a broad workflow is created and the different tasks are assigned. The work package also includes communication with all the key persons involved in the project.

Technical: The technical part of the project is divided into Data Acquisition, Processing, Analysis and Visualisation. In the Data Acquisition we collect reference data to be able to provide a quality assessment over our analysis. The Wi-Fi tracking is already done by TU Delft. The later steps will be more clearly defined as the project progresses.

Reporting: The final product of our research shall provide an analysis of use of space and recommendations for better use of space on the TU Delft campus to the client. This involves writing a scientific report that describes every step of the project in detail.

Public Relations: To present our findings a public presentation will be organised. The tasks involved in this work package involve inviting relevant people and organisations from the field. Additionally, an outreach to the media is important to make the project publicly known.

1.4. Project Logic Diagram

The project logic diagram shows an overview of our workflow including all the deliverables (framed content) and steps according to the WBS (descriptions at the edges), the full overview of the project logic diagram is attached in the appendix.

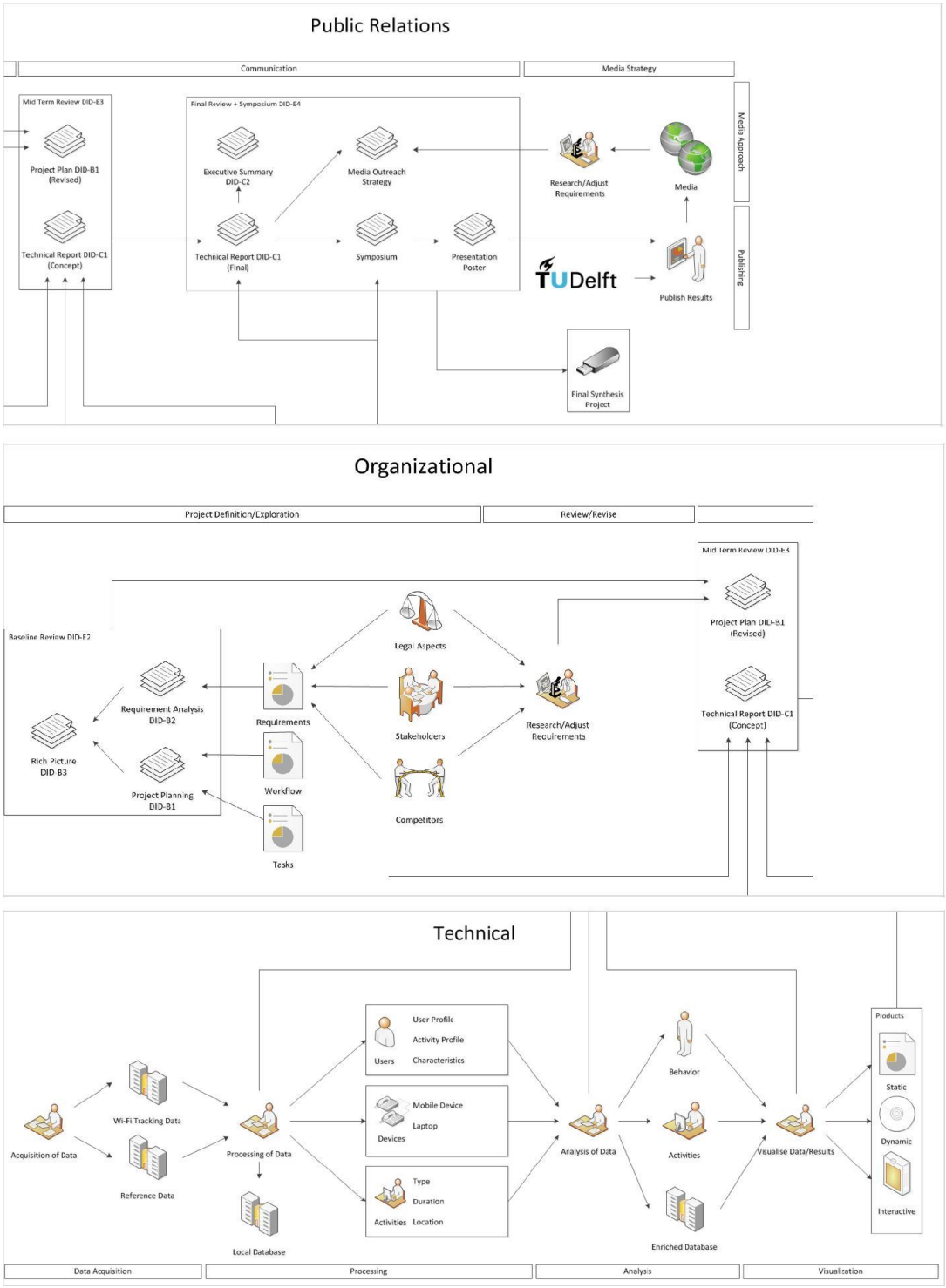


Figure 3: Project Logic Diagram

1.5. Schedule

An overview of the whole project including already scheduled meetings and deliverables is created in the form of a GANNT chart.

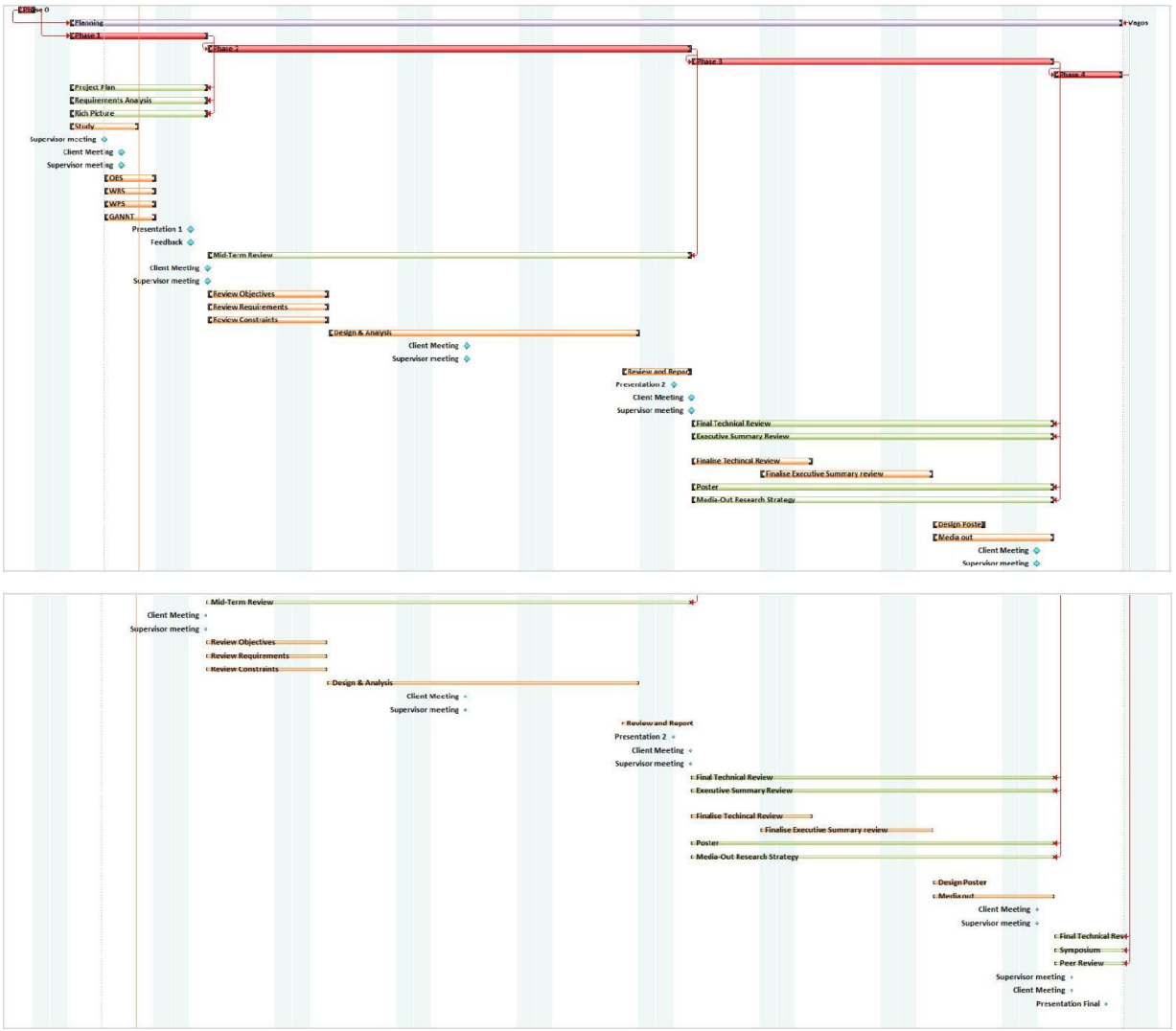


Figure 4: GANNT Chart

C. References

There are no sources in the current document.

D. Appendix

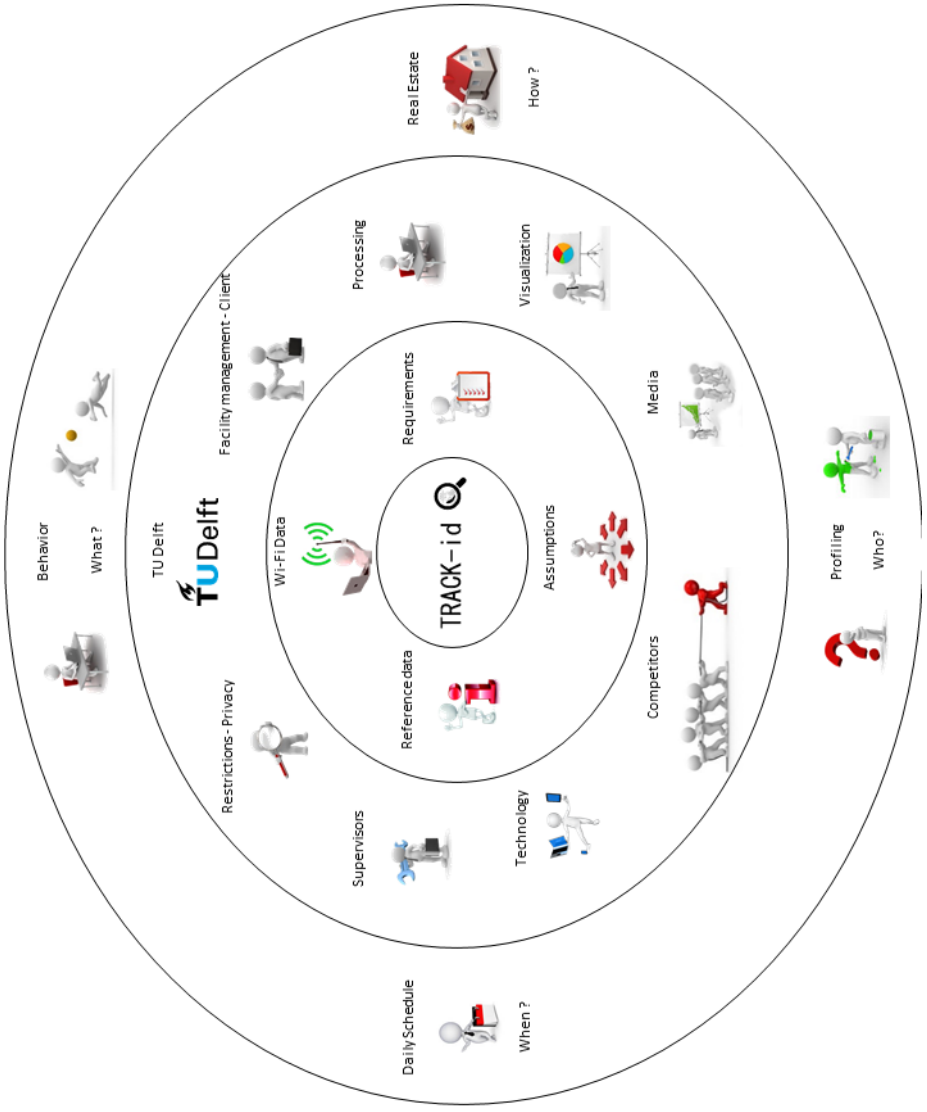
I. Rich Picture

II. Project Logic Diagram

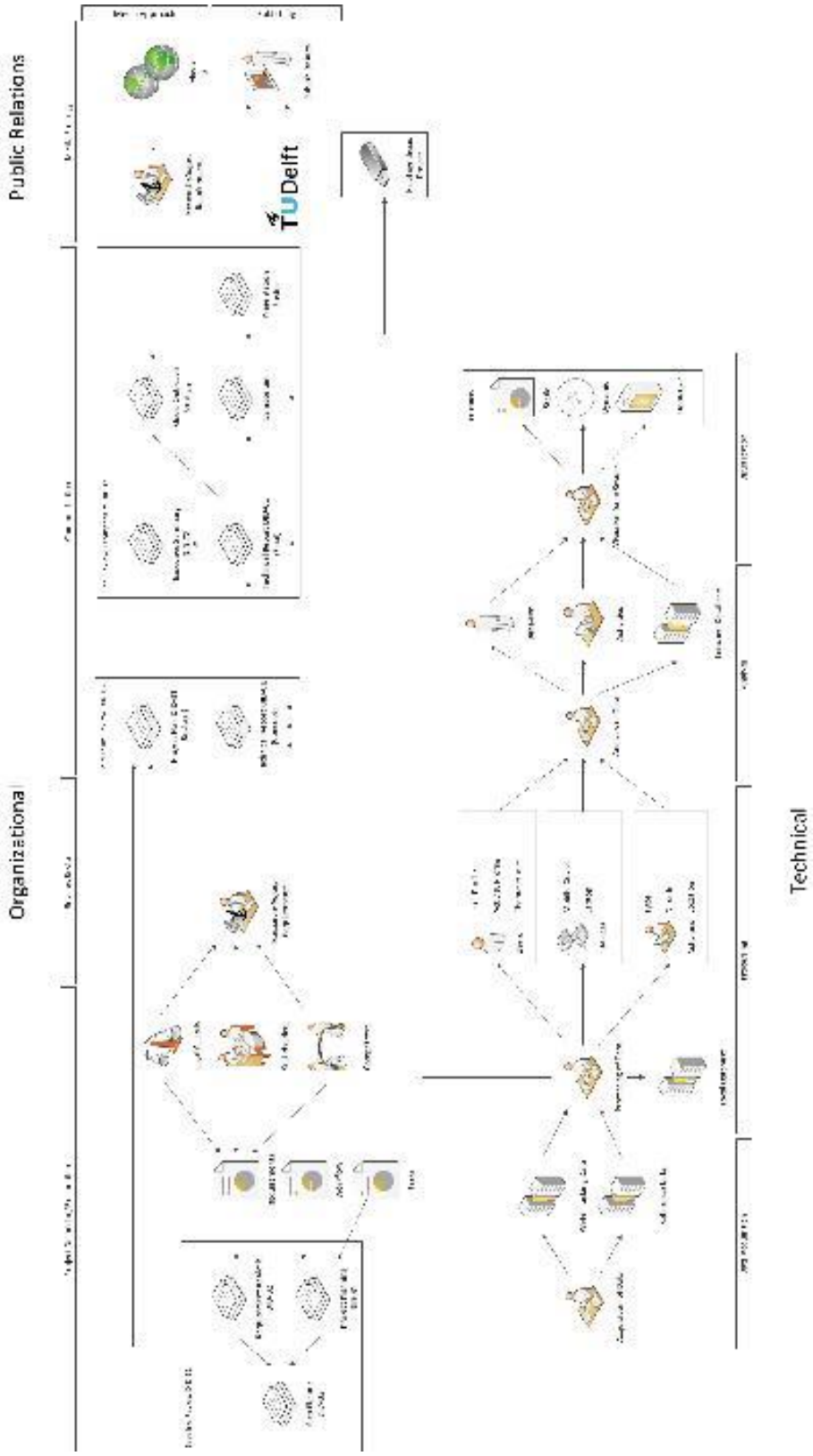
III. Requirement Analysis

IV. Requirement Tree

i. Rich Picture



ii. Project Logic Diagram



iii. Requirement Analysis

Requirements Specification

Stakeholder identification

TUDelft facility management department (beneficiary)
Activity organizer (beneficiary)
Students, employees, visitors and staff in TUDelft (potential beneficiary)
Big Brother (operator of the system)

Competitors

Groups doing statistical analysis using different technologies such as questionnaire Groups use a different tracking technology such as GPS or RFID

Top-level requirement

Use Wi-Fi tracking to determine the usage (how many, who are they, what they do) of the university during irregular hours.

Use Wi-Fi tracking to determine people's regular activity. (Perhaps)

Functional requirements (tasks)

1. Acquire Wi-Fi data.
2. Acquire reference information about access point locations and layout of faculties.
3. Relate access points to specific locations.
4. Which faculty.
5. Which floor.
6. Which room.
7. For how long.
8. Define if the device is a mobile phone or laptop.
9. By statistical analysis, identify a student, an employee, a visitor or a staff.
10. Assign user profile.
11. Speculate behaviour (lecture, lunching, studying in library, in laboratory, etc.).
12. Analyse people's behaviour during irregular hours (after opening hours).
13. Analyse participators of a known activity. From which faculty? What are they?
14. Validate the results.
15. Visualize the results using appropriate software.
16. Develop end-product for clients (report with recommendations).

Non-functional Requirements (criteria to assess)

1. The data can be obtained and a backup is created.
2. User profiles are correctly defined. (Faculty, what they are, etc.)
3. A well-defined sequence of behaviours that accords with someone's timetable.
4. People's behaviour accords with the observed Wi-Fi data.
5. The clients find the results are useful for future facility planning.

Technical requirements

1. PostSQL.
2. Programming.
3. Access points layout and technical parameters.
4. Participators have to carry at least one device with Wi-Fi function activated and connected to 'eduroam'.
5. Software for visualization.

Group requirements

1. Work as well-organized team.
2. Keep detailed logbook.
3. Keep up-to-date meeting minutes.
4. Regular contact with the clients.
5. High quality research content.

Known	Unknown
1. Wi-Fi data	1. Statistics about people recorded
2. Access points layout (maybe provided)	2. Accuracy of the dataset
3. University map	3. User profile
4. Four persona profiles	4. User behaviour
5. Activities agenda	5. Detailed regular activities
6. Activity information	6. Activity patterns (aiming group)

Constraints

User aspects

One person can have two or more devices with Wi-Fi function activated.

People walk around with their mobile phone while the laptops are left on desks. Some people don't turn on the Wi-Fi during the lecture or activity.

Some students or lecturers belong to faculty A may have lectures in another faculty time to time. (It is difficult to tag the user to which faculty he/she really belongs.)

Technology aspects

Only the RSSI at the moment when the device is connected to the access point will be recorded.

The access points are installed at the top of a room; it is hard to judge the actual floor (material of the floor affects).

Locations with different usages sharing the same access point. (for example: architecture studio and the restaurant of BK)

Some unmanned machines connected to access point automatically during irregular time; the records have to be found and avoided.

Privacy

Creating user profile and tagging participators are regulated by privacy (for now, the privacy is not in consideration).

It is possible to identify a specific individual during the research (In future work, it should be discussed into which extend the personal information can be intervened).

Assumptions

Assumptions will be made when speculating an activity. (For example, the behaviour of people who spend more than half an hour in library, is studying.)

The floor in the database is correct.

MoSCow rules

The MoSCoW technique is used by analysts and stakeholders for prioritizing requirements in a collaborative fashion (MUST SHOULD, COULD AND WON'T) (business analyst, 2013).

MUST describes a requirement that has to be satisfied, so that the process can be carried out without failure. In other words, the MUST requirements are the killer requirements.

SHOULD refers to a priority requirement that should be completed if possible within the time frame. COULD is a desirable requirement that will optimize the project if covered within the time allowed. WON'T is the requirement that is agreed among the stakeholders and developers to be contained in the future development and will not be completed this time.

1. Acquire data from database. 2. Access points layout and information. 3. Map of the campus. 4. Floor plans of buildings. 5. Relate access points to specific locations. 6. Layout of each faculty. 7. Pre-process data. 8. Make assumption about people. 9. Define user profile. 10. Analyse behaviour during irregular time. 11. Validate the results. 12. Discuss about privacy. 13. Deliver report with recommendations.	MUST
1. Make assumption about regular behaviour. 2. Visualization of the results. 3. Create maps to show the findings. 4. Discuss about the constraints. 5. Discuss about privacy issues.	SHOULD
1. Try to solve privacy problems (e.g. only use data from volunteers). 2. Model for people's behaviour prediction. 3. Determine regular activities based on assumed behaviour.	COULD
1. Improve technologies to improve data accuracy. 2. Develop end-products (applications).	WON'T

Geographical area concerned

All areas covered by TU-delft ‘eduroam’ signal; at a specific room level.

Software (organizational and technical)

Organizational
Google drive (share and exchange files)

Team-up (manage the schedule of each team member) Trello (manage workload)

Technical

pgAdmin
Python

iv. Requirement Tree