­

|  |
| --- |
|  |
| Research design for sensor database  Fontys IT & NovoData |
| |  |  |  | | --- | --- | --- | | Burak Basal, Martin Frenken, Nam Le, Hugo Meur, Jara van de Mortel, Tim Oomen, Djinn Opten, Jorrit Overeem, Hai Tran | 3/17/21 | Data Driven Business Lab | |

Content

[1 Introduction 4](#_Toc66806138)

[2 About Novo Data 4](#_Toc66806139)

[2.1 Meet the team 5](#_Toc66806140)

[2.2 Client background 6](#_Toc66806141)

[2.3 Problem description 7](#_Toc66806142)

[2.4 Expected result 7](#_Toc66806143)

[2.5 Conditions 8](#_Toc66806144)

[3 Constraints 9](#_Toc66806145)

[3.1 Scope 9](#_Toc66806146)

[3.2 Limitations 10](#_Toc66806147)

[4 Project method 11](#_Toc66806148)

[4.1 Iteration 11](#_Toc66806149)

[4.2 Why scrum? 12](#_Toc66806150)

[4.3 Tools 12](#_Toc66806151)

[4.3.1 Azure 12](#_Toc66806152)

[4.3.2 GitLab 13](#_Toc66806153)

[4.3.3 Docker 13](#_Toc66806154)

[4.4 Applied research 15](#_Toc66806155)

[4.4.1 Activities and strategies 16](#_Toc66806156)

[4.4.2 Long term planning 17](#_Toc66806157)

[5 Implementation 19](#_Toc66806158)

[5.1 Use of azure environment 19](#_Toc66806159)

[5.2 Data collection 20](#_Toc66806160)

[6 Project risks 21](#_Toc66806161)

[7 Project organization 23](#_Toc66806162)

[7.1 Agreements 23](#_Toc66806163)

[7.2 Roles within the organization 23](#_Toc66806164)

[8 Sources 26](#_Toc66806165)

# Introduction

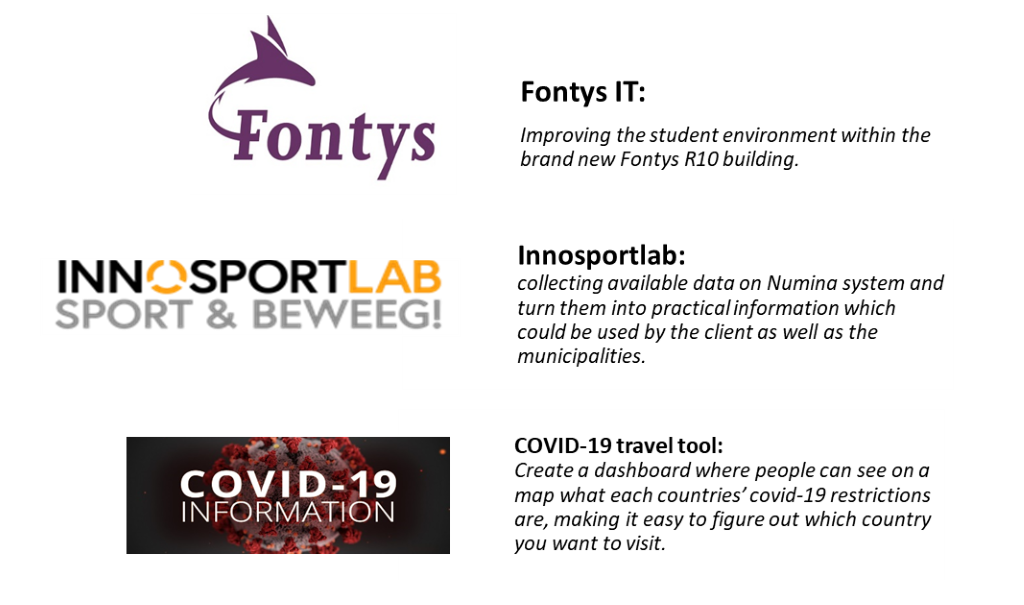
With new technologies, buildings can become more sophisticated. One of those buildings is R10 from Fontys University of Applied Sciences in Eindhoven, which has extra features like sensors to measure light, CO2-emissions and WiFi utilization. The project group Novo Data will combine these datasets with external data, to make a database with cleaned information.

The project will be executed between February and July 2021 in an online environment, due to the pandemic. We want to thank our process and content coach in advance, as well as our contact person from Fontys IT.

This document presents the project plan, which contains different topics. Chapter 2 gives some background information about Novo Data. Chapter 3 and 4 will mention information about Fontys IT and the project itself. The last chapters are about the working method, planning, risks and organization of the project.

# About Novo Data

Within the minor Data driven Business lab, Novo Data is one of the three participating companies. Novo Data consists of nine group members with different backgrounds, like ICT, Logistics Engineering and Industrial Engineering & Management.



## Meet the team

|  |  |  |
| --- | --- | --- |
| **Tim Oomen** | **Burak Başal** | **Martin Frenken** |
| Background in industrial engineering and management. Has a broad vision and experience with statistics and process optimalisation. | Background in trend research and concept creation, can provide insights into useful trends and translate them into new project concepts or improvements. | Experience in software engineering, project management and web engineering / networking. |
| **Jorrit** | **Jara** | **Hugo** |
| Background in programming and statistics. Has experience with visualization tools like PowerBI and Data analytics with R | Background in Logistics, experience with projects and problem solving. Has some knowledge about statistics, Power BI and programming in R. | Experience in software engineering and some background in statistics.  Has some knowledge in user experience. |
| **Hai** | **Nam** | **Djinn** |
| ICT & Software engineering, experience with software designer, front-end web developer. | Experience with software engineering, web engineering and data science. | Background in International Business and Aviation. Can carry out research and provide advice as an analyst. |

## Client background

The client is Fontys University of Applied Sciences in Eindhoven, which is a university of applied sciences in the Netherlands. The overall mission of Fontys is to offer inspiring, challenging and excellent higher education and to execute applied research that is meaningful for society. (Fontys, 2021) According to Fontys Hogescholen (2017), one of the actions for Fontys IT in Eindhoven is to use new technologies in new buildings to gain information about the usage of spaces/rooms, so that better choices can be made and to adjust the buildings to educational methods.

The new connected building R10 was opened on 24 August 2020 (see Figure 1). R10 was built as an activity-oriented working and learning environment. There are a lot of workspace types to fit the activities that take place. R10 has three different departments: FHICT, FPH Physiotherapy and Pulsed. R10 is equipped by a lot of sensors, so that the number of people in the building, the temperature choice for every room and usage of electricity and WiFi can be traced back. (Theeuwen, 2020) All this data must be anonymous and stay in the law with the GRPD.



Figure 1: R10 building

The stakeholder for Fontys University of Applied Sciences is Rens van der Vorst, who is the head of IT innovation and takes into account the impact of technology on society.

## Problem description

R10 is a new building with sophisticated sensors. These sensors have already been used for measuring and saving data in 2020, but nothing has been done with this data yet. Fontys IT wants to use this data for finding patterns that help to improve the education. Both internal and external data are needed to find correlations and influences, so the goal of Fontys IT cannot be achieved by only using the data of the sensors. Moreover, the collected data needs to be transformed to information for finding the correlations and influences.

## Expected result

As mentioned before, Fontys IT wants to use the data for finding patterns that help to improve the education. The main question of the whole project is as follows:

*“How can Fontys IT use data about buildings (sensors), occupancy and external data to find patterns that help to improve the education?”*

As this project will take more time than one semester, Novo Data will not be able to finish it. Therefore a different main questions is made for the start of the project, which is as follows:

*“How can R10 data and external data be used to obtain information for Fontys IT?”*

The task of Novo Data in this project is to start with collecting and combining useful internal and external data. As a result, the available data can be shown in a structured way in a database. If Novo Data has some time left, they can also give advice about which data should be analysed by using a dashboard. The steps that will be fulfilled by this group are mentioned below. Different questions are written down in the sequence of the execution:

1. What kind of available data from R10 and external sources is needed?
   * Search for data in the sensors.
   * Do a literature study about the latest possibilities in modern (smart) buildings.
   * Search for external data sources.
2. To what extent is this data reliable?
   * Check if the responses are in the right scales. Integrity of the data.
   * Make sure that the data can be used for improving the education.
3. How can this data be collected?
   * Find out how to get data from different sources on the Azure environment.
   * Find out what frequency for the data is needed
   * Find out how to connect different data sources with each other
4. How to store this data in databases?
   * Find out what tools for storing data are needed, and what architecture is feasible.
   * Take into account the transferability: how to scale up the project.

Novo Data will start this project with a small scale prototype, which will be explained in section 3.1. Next groups can continue the project by scaling up the prototype, making a cost overview, and applying machine learning theory to predict possible outcomes of variables based on certain patterns.

## Conditions

The main condition of the project is that it should be transferable, as Novo Data can only work on the project for five months. The new group should be able to use everything Novo Data has done already, including the programming code. This will be done by using Gitlab and Docker. More information about this will be given in section 4.3.

# Constraints

The constraints are divided into “scope” and “limitations". The scope mentions the elements that are included or not. The limitations are internal and external factors that have a negative influence on the research for the project.

## Scope

The scope dives into elements that are (not) included. Section 2.4 already mentioned the end deliverables from Novo Data, although it is difficult to estimate the tasks that can be done in one semester. That is why Figure 2 is made. It shows four squares, based on the Moscow theory. (Korolev, 2020) The green square shows the tasks Novo Data will complete at the end of the semester, in July 2021. This result can be achieved by giving answers to the research questions mentioned in section 2.4. When there is some time left, Novo Data can start with the yellow and orange squares. The red square gives an overview of things that will not be included. However, this only applies to the part Novo Data is doing for the project. Future project groups can choose to complete the yellow and orange tasks, when this is not yet done by Novo Data. They can also extent the project by deleting or changing the content of the red square.



Figure 2: Scope

Novo Data focuses on only one floor for the small database prototype, so that not all sensors have to be taken into account and the amount of data can be limited. The fourth floor consists of open ICT labs, working spaces and small speaking rooms. (Fontys ICT, 2020). Therefore it has different types of rooms, but not too many. Data for the fourth floor is probably more reliable, because nobody has to walk through this floor to go to another floor. As a result, the fourth floor will only be visited by people who have to be there for a specific reason. Because of these two reasons, Novo Data has decided to start the project with the fourth floor.

## Limitations

First of all, the principal limitation for this project is the pandemic of the corona virus: Novo Data can collect the current data of R10, but this data is not very representative for the normal usage of this building. Nevertheless, the previous data have been hosted in the database of ICT.

Also, Novo Data is limited by the good hardware function of the sensors: it has to check if the responses are in the right scales, otherwise the results in the statistics can be wrong.

# Project method

Scrum will be used for the project method. This is a framework that helps teams to work together. Much like a rugby team (where it gets its name) training for the big game. Scrum encourages teams to learn through experiences, self-organize while working on a problem, and reflect on their wins and losses to continuously improve.

These experiences are called sprints, where the team tries to accomplish a sprint goal. The sprint goal can be anything, for example creating a feasible project plan. If the team experiences any issues or delays during the sprint, they can reflect on that during the daily stand-ups, sprint reviews and retrospectives. The differences between these elements will be explained later on in this chapter.

Before a team can get started to work on their project, a sprint backlog is made. If a team has set their goal for the coming sprint, then they will discuss the different use cases and how long they think it will take to accomplish a use case. These use cases will be part of the backlog. (Atlassian, z.d.). A use case is a functionality that will be used in the product.

## Iteration

The period of a sprint depends on the project. Two weeks are the absolute minimum to work efficient with the scrum method.

During a sprint you have an implementation phase in which you reflect daily on the execution of the tasks within a use case. This reflection happens during the daily stand-up. This is the small iteration circle (see figure 3). The large circle indicates the iteration for the entire sprint, in which at the end a part of the product is delivered to the customer.

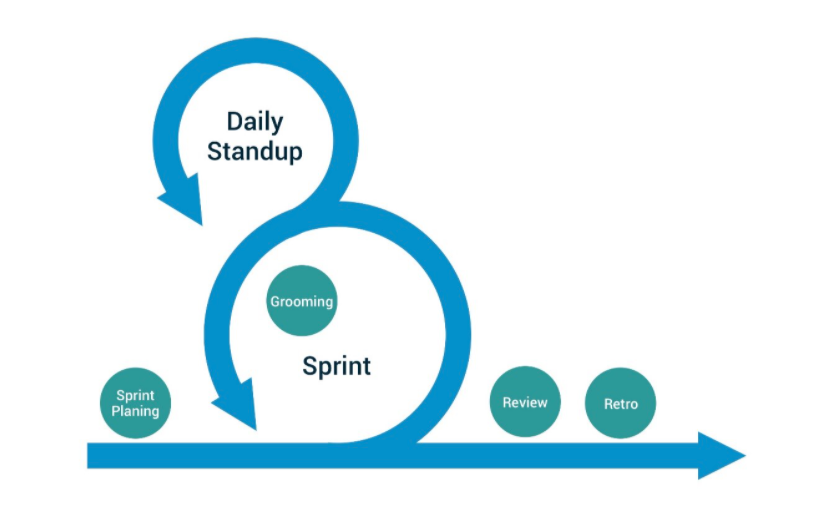


Figure 3: Scrum overview

Every daily iteration starts with a daily stand-up where a short meeting (max 15 minutes) where the team discusses their progress on the individual task. The scrum master asks his team the following questions individually:

* What have you accomplished yesterday?
* What task are you going to do today?
* Is there any help needed from other team members?

At the end of each sprint the team discusses their progress. The progress of product development there is a sprint review. Herein the team shows their product of the sprintgoal to the client and discuss the requirements of the next sprint goal. In the retrospective, the team reflects on the process as a team. This means that the team reflects on learning goals, what issues did slow down the process and how to solve them, but also to reflect on the gains.

## Why scrum?

An important reason for using scrum is that the project itself is quite large and doesn’t have a clear final goal. Novo Data has to think of usecases that can be applicable on the data and discuss these with the client. Some of these usecases can change during development and therefore Novo Data thinks scrum is suitable for this project.

## Tools

Several tools are needed to communicate and share things with team members for a successful project. For the communication, Teams and Outlook are used. Novo Data will use the cloud to store the data and to expose the dashboard on a website. Finally, to share the software product and the environment, Docker and Gitlab will be used.

### Azure

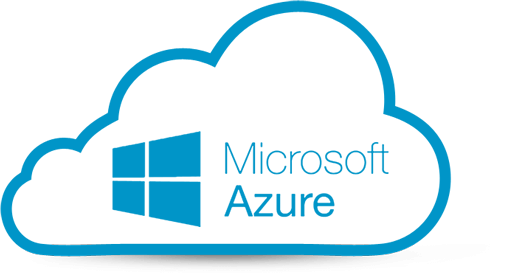


Figure 4 Logo Azure

Azure is a cloud computing platform. The cloud computing is a solution that provides everything to install applications. It can be used to store databases and to interact with the external API’s. Azure also provides computing solutions to interact with data in the goal to produce information. For example, algorithms can be hosted that will use the data to transform this in information in relation of KPI’s. Finally, it can also be used to host some dashboard or application.

### GitLab

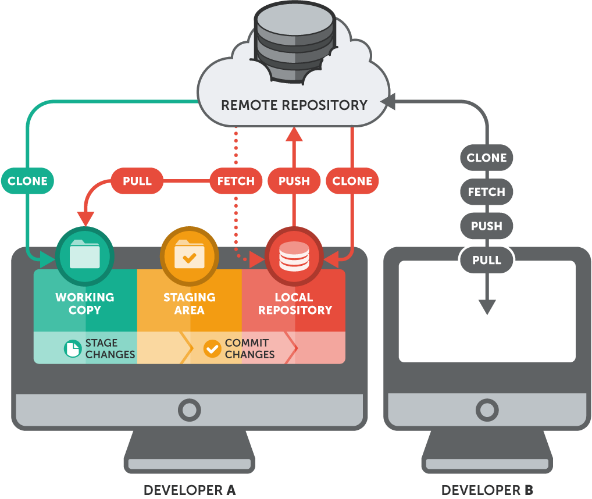


Figure 5 Overview Git

GitLab is a version control system that enables you to share and revert files, and to apply code review.

This tool is very efficient to avoid issues due to compatibility between our developments. Nevertheless, the conflict may be spawned to incompatibility of the environment installation. That is why Novo Data wants to use docker.

### Docker

Docker enables people to use the same environment. Figure 6 shows Docker can use one image, for example Python, to simulate one environment. When Novo Data has the same configuration file for Docker – which is shared by a Gitlab repository – Docker will pop the same container: it is like working in the same computer. So, with the combination of Gitlab and Docker, no issues due to the combability between environments will occur.

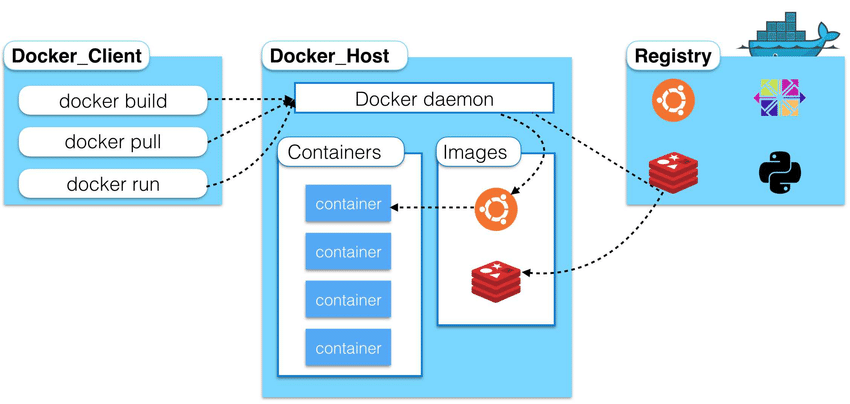


Figure 6: Overview Docker

## Applied research

A framework is needed to do applied research effectively. The DOT framework is chosen for this project, as it is related to ICT: it makes it easier to trace back the added value to the project. (Van Gennip, 2021) This framework has five different strategies, namely Library, Field, Lab, Showroom and Workshop. Figure 7 shows the corresponding activities for those strategies. (Heck, 2020)

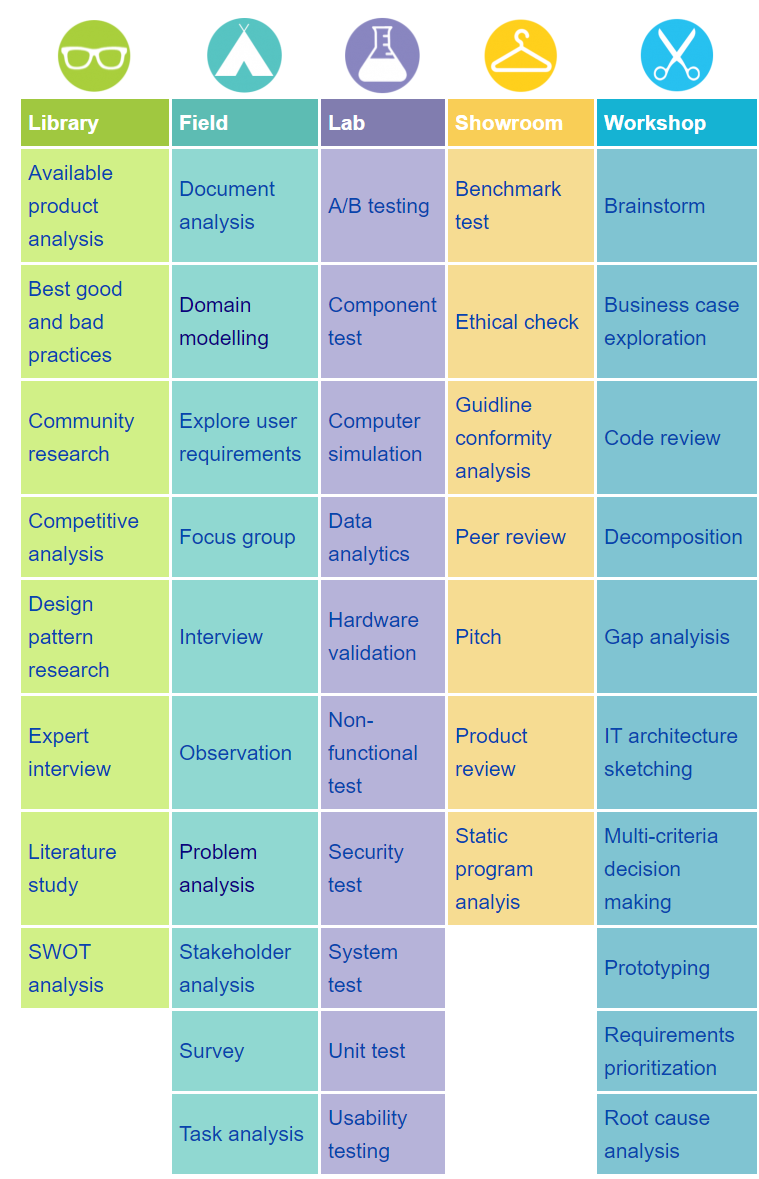
[](http://ictresearchmethods.nl/Methods)

Figure 7: Overview of DOT framework

### Activities and strategies

It is useful to know upfront what activities have to be done, so that it is easier to make a planning. (ICT research methods, 2018)

|  |  |  |  |
| --- | --- | --- | --- |
| **Research question** | **Research strategies** (CMD METHODS, 2020) | | |
| Sub questions |  | | |
| Sub question | Action | Strategy and activity | Stepping stones |
| What kind of available data from R10 and external sources do we need? | Search for data in the sensors. | Field research: interview + explore user requirements | Research report + interview report |
| Literature study about the latest possibilities in modern (smart) buildings. | Library research: literature study | Research report |
| To what extent is this data reliable? | Search for external data sources. | Library research: literature study + available product analysis | Research report |
| To what extent is this data reliable? | Check if the responses are in the right scales. | Lab research + showroom: system testing or non-functional test + prototype | Analytics report + prototype |
| How can this data be collected? | Find out how to get data from different sources on the A­zure environment. | Library research: available product analysis | Research report |
| Find out what frequency we need for the data | Lab research: component testing | Analytics report |
| Find out how to connect different data sources with each other | Workshop + showroom: IT architecture sketching + prototyping | Prototype |
| How to store this data in databases? | Find out what tools we need to store the data | Library research: best good and bad practices | Research report |
| Find out what architecture is feasible. | Lab research: IT architecture sketching | Analytics report |
| Take into account the transferability: how to scale up the project. | Lab research: system testing + workshop: prototyping | Prototype |
| Find out if it works | Showroom: product review | Prototype |

### Long term planning

The activities mentioned above are combined with the Scrum method in a Gantt chart planning. This planning exists of sprints that take two weeks. Sprints of three or four weeks would be too long, as the project only takes 18 weeks. Moreover, sprints of four weeks have less deadlines so that the pressure to perform would be weak. Every sprint focuses on one of the research questions and corresponding action(s). Figure 8 shows the tasks and approximate deadlines of every task and sprint.



Figure : Gantt chart planning

# Implementation

This chapter explains how Azure will be used as a service environment for the research, and how a dashboard will be applied to visualize the data. Section 5.1 explains the software architecture, and why a certain service is needed Section 5.2 describes why and which data is going to be collected. Depending on the cost and the amount of data, the dashboard should be refreshed each week.

## Use of azure environment

The general idea of Fontys IT is to use Azure as a cloud service for data collection. Azure is a cloud computing environment from Microsoft for different internet services. Azure Kubernetes services will be used to collect the data from external data sources and do a first small cleaning for sensors and WiFi controllers and store it into database. Novo Data wants to be able to know when and what type of data is coming. That is why it wants to create one container for each source. With this system, Kubernetes will be able to schedule independently each launch of program to import the data. After that, the data cleaning and the formatting will be done in a second step with python containers. We will store the clean and ready to use data and database 2.

The detailed overview of the steps is mentioned below. This is based on Figure 9.

1. The first step is to find external API’s that can be compared with the data of the Fontys IT database.
2. These different datasets will be combined by using Kubernetes services. This service makes it possible to clean the raw data, for example outliers and negative figures.
3. All remained data will be stored in the Azure database 1.
4. To be able to compare different variables in a dashboard, the data should be divided into different containers and they should be set at the same scales. For example, the external data should be based on hours, if the sensor data is based on hours as well. This will also be done in Python by Kubernetes, as this service makes it possible to load datasets individually.
5. These datasets will be stored in the Azure database 2.
6. The client wants to have the end result in Power BI, so that the data will be turned into information. Therefore, Novo Data has to make a link between the clean database and Power BI.

Novo Data will use Python in the containers to clean and arrange the data easily. Python is one of the more common language and it is powerful with Big Data. Moreover, a lot of libraries are available, which could help with statistics and visualization.

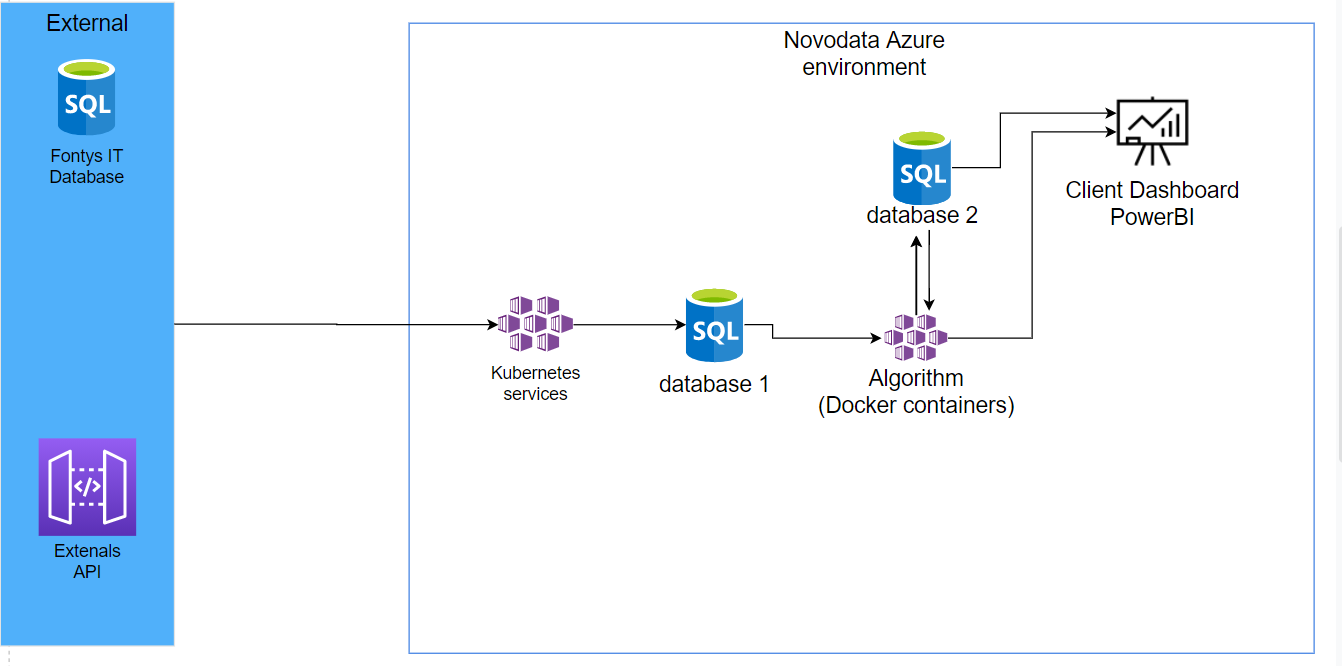


Figure 9: Architecture

## Data collection

May be we don’t need all the data from the sensors but just a small part to get a first insight and not too exceed the scope of the project. Therefore we want to collection only the data from the sensors and WIFI controllers that are on the first floor.

For the external data we want to collect data about the weather. The KNMI(dutch meteorological institute). So for historical weather data we have to look for other sources. The weather could have an effect on how many students are present during the day.

We can also collect data of the humidity within the building. Maybe we can access the Wi-Fi controllers to see if at some point students leave the building or room when the humidity values are too high.

# Project risks

The risks are mentioned in the table below. The probability is based on the following scale:

* 1: not very likely. The chance that it does not happen is bigger than the chance that it happens.
* 2: possible. The chance that it happens is approximately as big as the chance that it does not happen.
* 3; very likely. The chance that it happens is bigger than the chance that it does not happen.

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Action** | **Probability (1-3)** | **Impact (1-3)** |
| A project member gets ill for a few days or weeks. (general) | Students of other groups can take over some tasks, as Novo Data exists of a group of nine students. | 2 | 2 |
| A group member struggles to learn a software or programming language / platform / control system (general) | A member of the team who knows more about the programming language assists the other member, possibly with a workshop or pair programming | 3 | 1 |
| Data gets exposed unwillingly to outside users (technical) | We will make sure to secure our API's so that outside will not get access to our data. | 1 | 3 |
| The hosting service of the project application runs out of cloud credits (technical) | We will carefully monitor cloud credit expenditure and supervise new changes that are being made in the cloud environment. | 1 | 2 |
| There is not enough data available | Getting extra data from the sensors is difficult due to the lockdown. However, external data is always available. | 2 | 2 |
| The available data is unreliable. | It does not make a difference in the way of working (collecting the data and making choices in the variables to take into consideration). Therefore, certain models/dashboards could still be used when the right data is updated. However, using unreliable data has impact on the conclusions when analysing the data. | 1 | 3 |
| Limited by the good hardware function of the sensors. | We have to check if the responses are in the right scales. If we don’t check the responses of the sensors, we may have false result in the statistics. | 2 | 2 |

# Project organization

This chapter contains the contact moments and agreements between the project team, the client and the project supervisors. In addition, this chapter includes the contact information of all stakeholders within this project.

## Agreements

The agreements made are mentioned below:

* Every group member needs to be available from 9:00 to 16:00, from Monday till Friday;
* Stand up meeting every Monday, Tuesday, Thursday and Friday at 9:30 with NovoData;
* Fixed lunch break every day between 12:00 and 13:00;
* Daily standup from 9:00 to 9.15 with the Fontys IT group;
* The progress of the project will be discussed after every sprint. This meeting will take place on the last Friday of the sprint between 14:00 and 15:00;
* Novo Data will meet with Rens van der Vorst every Friday at 10:30, and with Jeffrey every Thursday at 9:00. During these meetings, the progress will be discussed and questions from both sides can be asked.

Moreover, certain deadlines should be taken into account to fulfil the project:

* The project plan and infographic should be finished at the end of week 5. The presentations of these deliverables will also be in this week.
* Around week 12 there will be a midterm review. Because of this, Novo Data will start preparing specific deliverables for this in week 9.
* Week 17: wrapping up projects.
* The project ends with an end presentation and project review in week 18/19.

## Roles within the organization

The following tables will explain the roles and contact details per person for this project.

|  |  |  |
| --- | --- | --- |
| **Who** | **Role** | **Description** |
| Jara | Note taker and planner / Scrum master | Making notes during the meetings, and planning tasks ahead. |
| Hugo | Developer | Developing the technical software. |
| Jorrit | Projectleader / Developer | Leading role during the project and developing software. He communicates with the client to make sure all requirements are in the project. |

Table 1: Project members

|  |  |  |  |
| --- | --- | --- | --- |
| **Who** | **Role** | **Mail** | **Phone** |
| Rens van der Vorst | the head of IT innovation, contact person and client | r.vandervorst@fontys.nl | 0885079817  06 50 24 20 71 |
| Mike Baarslag | Dienst IT, knowledge about Fontys database | m.baarslag@fontys.nl | 0885075910  06 50 24 20 88 |
| Björn van Veldhoven | Programmeur Dienst IT, can help with Azure | b.vanveldhoven-ab@fontys.nl | 0885078201  06 57 31 23 24 |
| Peter Swinkels | Advisor, can help with Azure | p.swinkels@fontys.nl | 0885078174  06 53 23 04 40 |
| Ron Bertrams | Building management system operator, can answer questions about Excel files | r.bertrams@fontys.nl |  |
| Martijn | 12CU specialist, contactpersoon van Rens | [support@12cu.com](mailto:support@12cu.com) |  |

Table 2: Contact details clients and stakeholders

|  |  |  |  |
| --- | --- | --- | --- |
| Who | Role | Mail | Phone |
| Jeffrey Cornelissen | Content coach | jeffrey.cornelissen@fontys.nl | 0885074941  06 20 58 52 00 |
| Bart van Gennip | Process coach | b.vangennip@fontys.nl | 0885078966  06 57 87 56 54 |

Table 3: Contact details coaches

|  |  |  |
| --- | --- | --- |
| **Who** | **Mail** | **Phone** |
| Jara | j.vandemortel@student.fontys.nl | +31651053322 |
| Jorrit | j.overeem@student.fontys.nl | +31610502559 |
| Hugo | h.meur@student.fontys.nl | +33646118744 |

Table 4: Contact details project members

# Sources

Atlassian. (z.d.). *What is Scrum?* Geraadpleegd op 2 maart 2021, van <https://www.atlassian.com/agile/scrum#:%7E:text=Scrum%20is%20a%20framework%20that,and%20losses%20to%20continuously%20improve>.

Fontys. (2021). *Over Fontys*. Geraadpleegd op 15 maart 2021, van <https://fontys.nl/Over-Fontys.htm>

Fontys Hogescholen. (2017). *Fontys Focus Informatievoorziening*. Geraadpleegd van https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjz2rqyqrLvAhXHu6QKHXfdALoQFjAAegQIARAD&url=https%3A%2F%2Ffontys.nl%2FOver-Fontys%2FNieuws-tonen-op%2FInformatievoorziening.htm&usg=AOvVaw2UpW\_1II0al6DJ83baq7Et

Fontys ICT. (2020). *We are open!* Geraadpleegd van https://www.partnersfontysict.nl/site/assets/files/5416/wegwijzer\_r10\_fhict-en\_interactief\_comp.pdf

Heck, P. (2020, 20 november). *ICT Research Methods for Machine Learning Engineering*. Geraadpleegd op 9 maart 2021, van https://fontysblogt.nl/ict-research-methods-for-machine-learning-engineering/

ICT research methods. (2018, 9 februari). *Methods - ICT research methods*. Geraadpleegd op 9 maart 2021, van <http://ictresearchmethods.nl/Methods>

Korolev, S. (2020, 6 maart). *MoSCoW Method: How to make the best of prioritization*. Geraadpleegd op 21-03-10, van <https://railsware.com/blog/moscow-prioritization/#What_is_MoSCoW>

Theeuwen, M. (2020, 24 augustus). *Fontys opent “duurzaam pareltje” R10 op campus Eindhoven*. Geraadpleegd op 4 maart 2021, van https://www.ed.nl/eindhoven/fontys-opent-duurzaam-pareltje-r10-op-campus-eindhoven~a4938cc1/?referrer=https%3A%2F%2Fwww.google.com%2F

Van Gennip, B. (2021, februari). *The quality of projects*. Workshop gepresenteerd bij Workshop Applied Research, Eindhoven, Nederland. Geraadpleegd van <https://stichtingfontys.sharepoint.com/:p:/r/sites/GastLecturesWorkshopsTQ4.1/_layouts/15/Doc.aspx?sourcedoc=%7B6F2E0DF6-A2B2-4D55-8247-978399B48DA3%7D&file=DDBL%20RnD%20Rigor%20DesignOrientedResearch%202021vj.pptx&action=edit&mobileredirect=true&wdPreviousSession=ce77e2aa-c5f4-4720-8503-1f40d67f2b2d&wdOrigin=TEAMS-ELECTRON.teams.undefined>