Muntstroom PCP, Brussels Capital Region

Pre-Commercial Procurement (PCP) regarding
R&D of end-to-end solutions for monitoring
multi-faceted people flow

Market consultation report

The Muntstroom PCP Group:

STIB-MIVB (lead procurer)

CIRB – CIBG

Brussels Mobility

Parking.Brussels

Contact person:
Paul van Houtte, Paul.VanHoutte@mivb.brussels

With the support of:

innoviris.brussels
empowering research

European Assistance For Innovation Procurement
eafip

Disclaimer & Copyright

This document (including the accompanying annexes) has been drafted with the utmost care, but no guarantees are given regarding its soundness and/or completeness. Any errors or inaccuracies can be reported via muntstroom@stib-mivb.brussels.

This report has been made available in English. A French and/or Dutch translation may be provided at request.

Muntstroom PCP Group is not responsible for the correct operation of any URL mentioned in this document, nor for the proper functioning of any used electronic platform (for example the Belgian Public Tender Bulletin or the EU survey system). Any problems encountered when using a URL and / or an electronic platform must be reported to the organisation that makes the URL or the electronic platform available. Problems with downloading and uploading (of documents) must also be reported via muntstroom@stib-mivb.brussels.

In conducting the Open Market Consultation and Open Client Dialogue, Muntstroom PCP Group benefits from assistance offered under the Eafip initiative[1], which is supported by the European Commission Directorate-General Communications Networks, Content & Technology (DG CONNECT).

Economic operators are being informed that any information regarding the setup and execution of both the procurement process and the execution of any contract/framework agreement as a result of the procurement process as well as public summaries of the results of the PCP project, including information about key R&D results attained and lessons learnt by the procurers during the PCP, can be shared after consultation with the respective R&D provider by the Muntstroom PCP group with(in) the context of the Eafip initiative and consequently can be analysed, (re-)used and published by the Eafip initiative. Details should not be disclosed that would hinder application of the law, would be contrary to the public interest, would harm the legitimate business interests of the R&D providers involved in the PCP or could distort fair competition between the participating R&D providers or others on the market.

[1] www.eafip.eu

Content

1 Purpose of this document

The Muntstroom PCP is currently in phase 0. See the figure below. It started with an extensive preparation by the public buyers. Subsequently two market consultations have taken place:

1. an Open Market Consultation (OMC) to provide the Muntstroom PCP Group with a comprehensive feedback from technology suppliers and
2. an Open Client Dialogue (OCD) to provide the Muntstroom PCP Group with a comprehensive feedback from potential end-users.

This document presents an anonymised report of the results of the two market consultations. It has been published via the project website[2] and/or EUSurvey portal on the date indicated in the timetable of the Market consultation document.

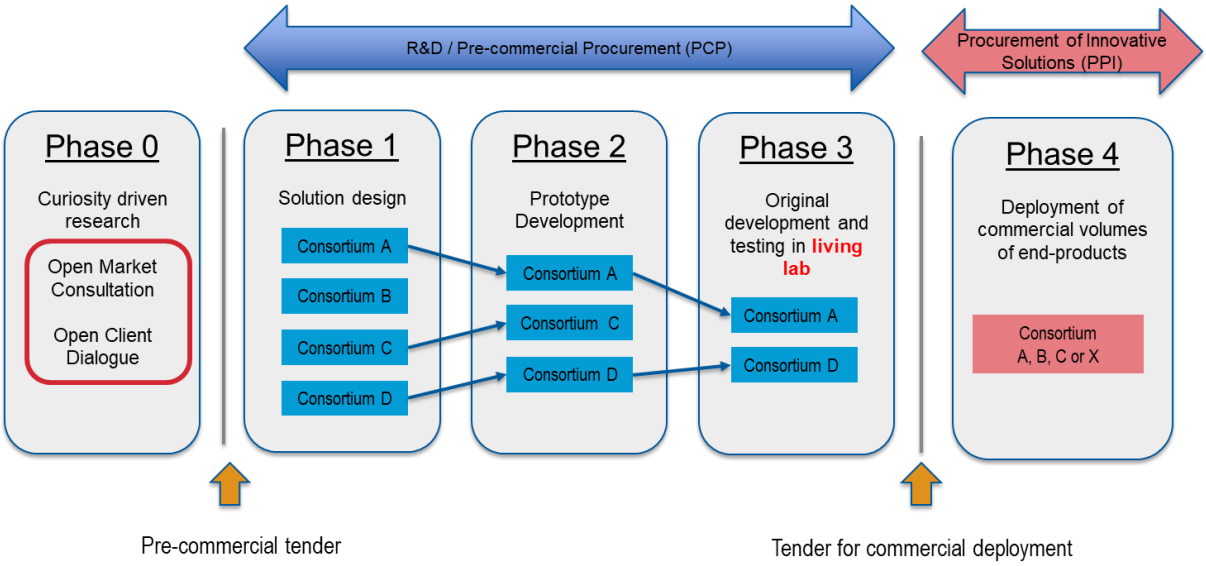

Figure 1: The place of the market consultion within the different phases of Innovation procurement

[2] www.muntstroom.brussels

2 Muntstroom PCP

The Muntstroom PCP is a pre-commercial procurement (PCP) that aims at developing and testing an integrated end-to-end solution for monitoring people flow. The project is a result of a project call by Innoviris[3] for Smart Mobility Living Labs in the Brussels-Capital Region (BCR).

The Muntstroom PCP group / public buyers

Four public partners submitted a joint proposal, being:

1. Public transport operator STIB-MIVB
2. Regional Informatics Centre CIRB-CIBG
3. Regional authority Brussels Mobility
4. Regional agency Parking Brussels

Assistance providers

The proposal was accepted by the Government of the Brussels-Capital Region in 2019. As a result, the project is partly funded and supported by Innoviris. The Muntstroom PCP has also been selected by the European Commission to receive support from the Eafip initiative[4].

Consultancy

Via the assistance providers, the Muntstroom PCP Group is supported by two consultancy firms: Equator Law and Corvers Procurement Services B.V.

Outreach group

For the outreach of the Muntstroom project, the public buyers work together with 4 business support organisations: EASME, Agoria, BECI and Hub.Brussels. These parties are helping with reaching out to Brussel's, Belgium's and Europe's most innovative companies. Additionally, they will help identifying the matchmaking needs and subsequently provide the necessary matchmaking services.

Lead procurer

The joint PCP will be conducted by STIB-MIVB as lead procurer in representation of the Muntstroom PCP Group under Belgian law.

Budget

The foreseen budget for co-financing the R&D is maximum € 500.000.

This budget does not include the possible future public procurement of the to be developed solutions (PPI), as a result of this PCP.

[3] The Brussels institute for encouraging scientific research and innovation

[4] European Assistance for Innovation Procurement initiative

3 The underlying means

3.1 Markt consultation document

As basis for the two market consultations one general Market consultation document (click here) was written by the public buyers. Because the Muntstroom project combines three innovative approaches[5], the Market consultation document was divided into six parts that could be read separately.

- Part 1: Introduction of the Muntstroom project
- Part 2: Desired solution and desired output
- Part 3: What is Pre-Commercial Procurement
- Part 4: Open Market Consultation
- Part 5: Open Client Dialogue
- Part 6: Follow up market consultation

Additional four annexes were available

- Annex 1) Scope of the project (click here)
- Annex 2) Use cases (click here)
- Annex 3) OMC-questionnaire on EUSurvey (click here)
- Annex 4) OCD-questionnaire on EUSurvey (click here)

3.2 Animated video's

Although the rules and background information of both the OCD and OMC are similar, as they are conducted in the context of the Muntstroom PCP, the questions addressed and feedback obtained differ due to the nature of the participants, their particular interest and roles.

Hence the public buyers made two animated videos that explained the process:

1. Muntstroom challenge for technology vendors: https://youtu.be/jUYMmYl4ebg
2. Muntstroom challenge for parties interested in People Flow-data: https://youtu.be/W816gbYeHDI

3.3 OMC-webinar

The Muntstroom PCP Group organized a digital information meeting about the project. The meeting was held via Microsoft Teams on the 19th of November 2020.

A total number of 106 registered and finally 62 people participated in the webinar.

The slides of the webinar can be found via this link (click here)

[5] 1) the desired end-to-end solution, 2) the PCP-approach and 3) market consultations to both a) the technology providers and b) to the potential users of the to-be-produced People Flow data.

The recording of the webinars can be found via the following links:

- OMC-video part 1: Introduction
- OMC-video part 2: Challenges from desk research
- OMC-video part 3: The PCP-approach
- OMC-video part 4: The process of the Open Market Consultation

In total 39 interested parties replied to the OMC-questionnaire.

3.4 OMC-questionnaire

Interested market parties were requested to fill out the OMC-questionnaire that was published on EUSurvey.

The questionnaire included 10 sections and a total of 49 questions (see table below).

The general findings and results are summarised in the following sections.

Section	Questions No.
1. Contact details	-
2. About you	1 - 5
3. Desired output and needs	6
4. Architecture	7 - 21
5. Use cases (required analytics)	22 - 26
6. Privacy & security	27 - 31
7. Data quality	32
8. Financial aspects	33 - 44
9. PCP challenges and complexity	45 - 48
10. Last remarks	49

3.5 OCD-webinar

The Muntstroom PCP Group organized a digital information meeting about the project. The meeting was held via Microsoft Teams on the 19th of November 2020.

A total number of 34 people registered and 31 people finally participated in the webinar.

The slides of the webinar can be found via this link (click here)

The recording of the webinars can be found via the following link:

- OCD-video part 1: Introduction
- OCD-video part 2: Our questions to you
- OCD-video part 3: The process of the Open Client Dialogue

In total 8 interested parties replied to the OCD-questionnaire.

3.6 OCD-questionnaire

Interested market parties were requested to fill out the OCD-questionnaire that was published on EUSurvey.

The questionnaire included 8 sections and a total of 27 questions (see table below).

Section	Questions No.
1. Contact details	-
2. Your interest in People Flow	1 - 2
3. Participation in the PCP	3 - 4
4. Use cases	5 - 7
5. Your needs	8 - 21
6. Value of data	22 - 24
7. Privacy and security	25 - 26
8. Last remarks	27

4 Summary results OMC-questionnaire

This chapter summarises the results of the Open Market Consultation. This market consultation was targeting companies that could provide (elements of) the desired integrated end-to-end solution to produce People Flow-data.

4.1 About the respondents

Most of the respondents work in SMEs and are located in Belgium.

The majority of companies are active in the field of analytics and analytics visualisation. Other respondents are active in the field of Cybersecurity, Mobility solutions and ICT, with a special focus on the physical security industry, computer vision, AI & deep learning services. Other have specified to be active: in data warehousing. data management, integration, IoT, Natural Language Processing, Open Data portal, Commercial Information Portal, IoT Accelerator, Data broker, PaaS for containerised workload, Data lake / Data warehouse, Low Code Rapid App Development, Aligned with standards alike OASC, Fiware. One is active in emergency mobile health app. One is active in Integration Services, Managed Services.

In total 39 interested parties replied to the OMC-questionnaire.

31 out of 39 are looking for a partners a matchmaking. The list with parties interested in matchmaking has been send to the members of the outreach group. In Q1/Q2 2021 they will organise matchmaking activities and the interested parties will be invited for this.

The OMC-questionnaire was available in English, French and Dutch. The language use in the open questions was distributed as follows.

	Respondents
English	35
French	2
Dutch	1
No written answers	1

Question 1. In what kind of company are you working?

	Answers
Small-sized company (less than 50 employees)	19
Medium-sized enterprise (between 50 and 249 employees)	7
Large enterprise (from 250 employees)	12
No Answer	1

Question 2. Where is your company based?

	Answers
Brussels Capital Region	8
Belgium	21
Europe	10
Outside Europe	0
No Answer	0

Question 3. In which field is your company active? (multiple answers possible).

	Answers
a. Data capture (incl. sensor technologies)	24
b. Data analytics	37
c. Data visualisation / services	33
d. Infrastructure (incl. backend system and front end system)	21
e. Telecommunication	6
f. Other ….	12
No Answer	0

Question 4. How do you envisage to be engaged in this PCP-tender procedure?

	Answers
a. Consortium (several technology vendors)	27
b. Main contractor with sub-contractor(s)	7
c. Sole contractor	0
d. I do not know yet	5
No Answer	0

Question 5. Are you looking for (a) partner(s) with complementary expertise to form a consortium for the purpose of submitting an offer on the Muntstroom PCP?

	Answers
Yes	31
No	8
No Answer	0

4.2 Desired output and needs

The Muntstroom PCP Group would like to develop and test a solution to visualise 24/7-people flow (counting, direction, speed), facilitate shared big data-analytics and make the People Flow-data available for a wide array of users. The content of the scope document will be addressed on the basis of several questions.

To describe the underlying needs, visualisations were made of the desired solution and the proposed data architecture. See Annex A of this document.

Question 6. Do you have any general recommendations or suggestions regarding these figures? (Open Question)

The general recommendations and suggestions received are summarised as follows:

- The output could be amended to “Applications tools”. (Routing is one good example mentioned)
- Maybe also add the integration of sensors trough an eco-system (at the device layers).
- In order to secure personal data, it could be an option to provide an additional processing layer between the sensor layer and the communication layer that converts personal data into anonymised data. Although this is not preferred, this option should be kept open during the design phase.
- We see the use and combination of different technologies as a key deliverable of this project. As such, the people flow data standard should be at the output & not on the input.
- As different technologies will be used as input, there will most likely be a need for a metadata reference store, which links the measurement technologies to the environment. Examples are already available within smart buildings / digital twins. These concepts could be reused and are missing in the architecture
- On the diagram a monitoring block could be added. This monitoring is important to manage the health of the different pieces of the pipeline.
- Still unclear what is meant with 'processing' (aggregation + quality check), even after reading Annex 1 (3.4.basic activities, and 3.5 preparing for analytics)
- Yes, considering that thermal images are always privacy-respectful, including periodical samples of those images being captured (e.g. every 30’) into the catalogue of open-source data before its processing could help to data-users/clients to figure out new data to be extracted from such images, to cocreate new uses for them.
- You could start from a central cloud infrastructure. Having the analytics closer to the sensor could help with privacy aspects
- I would also add the correlation process in the Aggregation+Processing box. This because data might come from different sources, different authorities, different levels of reliability,…

- The time dimension is missing from the diagram. You need to show that the user needs span from real-time monitoring to planning and forecasting
- Input data (sensors) should be more specific, because there is significant data quality difference once compared e.g. beacons with cameras. Regarding this you could create sensor categories for particular groups of data gathering needs.
- There are "proprietary" components doable with an open approach as well. No need to close that part.
- This architecture seems to suggest all sensor data is sent to the cloud and processed there. This could lead to inefficiency in performance (due to latency) as well as cost price (data communication and storage). Through using smart cameras e.g. the image processing / anonymization could already be done on the edge.
- Activity data collected in a Routing App could be a valuable additional data source, which seems not to have been considered here.
- You could add trough layer interfaces: a possibility to integrate the solutions into other components.
- The transversal "Privacy & Security-by-design" layer could also be used to allow dynamic external data to be included in all layers of the stack. The "meta" data would not need to be stored, it would just be used on-the-fly during the various processings, from capture to the end user. The aim is to use the freshest possible data.

About standards

- We normalise all the data sources in linked data through open standards as RML to facilitate organisation and inference via knowledge graphs.
- Rather than emphasizing on standards, you could put emphasis on the openness as the basic priority for device interconnectivity as standardisation keeps evolving. Interfaces should be based on open interconnections.
- Data collection can be done on open standards (e.g. RTSP, MPEG4, ...)
- Government should be the provider of which open standards to adhere to. If no standards exist, which is doubtful, then own standards need to be defined by government. The https://www.fiware.org/about-us/ could be an inspiration, or the OGC Sensor specification.

4.3 Architecture

4.3.1 Physical layer

People Flow-data standard

When you look into the field of walking, it becomes clear that no pedestrian is alike. To well analyse any situation it is thus important that the metrics of People Flow are well defined. The Muntstroom PCP Group therefore wants to use the Public Life Data Protocol[6] (PLDP, 2017) of Gehl Institute, that defines 3 levels of details for describing the category pedestrians.

Question 7. Can you indicate till which level of detail is it possible to capture People Flow-characteristics? (Via your technical solution or in general) Pedestrian (in general)?

	Answers
a. The subcategories (Walking, Running, Supported, Carried, Rolling)	12
b. Advance content (more details within the subcategories)	10
c. No specific metrics	1
d. Other, namely	9
No Answer	7

If other, please indicate

- The subcategories will be possible to detect (via vision). The advanced content is a field of research, which does not mean it is not feasible either, but this will require a different mindset from the tendering authority.
- The subcategories mentioned (Walking, Running, Supported, Carried, Rolling) plus Population, Density and Direction, as well as proximity to points of interest, points of activity
- "Walking" and "running" based on speed is possible to capture. "Supported/carried/rolling" is already much harder, if not impossible, to capture.
- Yes for "Walking" and "Running". Certain other categories can be considered depending on the camera viewpoint via AI (stroller, prams, wheelchair...) as well as probably certain scenarios of "carried" people.
- It is possible to obtain all these sub-categories and more on a pedestrian, but the question may arise as to the relevance of working only on this individual scale. This "Individuation" is namely very interesting in itself but should -in our view- be complemented by the study of crowds. In other words, matching microscopic and macroscopic visions by observing both the individual and the crowd/flow of which he or she is a part.
- Even when using thermal cameras, pedestrians, bikers, scooters, skates, segways… can be differentiated and classified.

[6] https://gehlpeople.com/tools/public-life-data-protocol-beta/

- Level (a) (identifying the subcategories) is possible via computer vision. Other things are possible too, and these are very relevant to the project, e.g.:
    - mapping of detected people to groundplans or floorplans. This will need some sort of calibration, in which pixel positions in camera images are mapped towards metric units (i.e. coordinates on a map)
    - assessing other characteristics, such as direction and speed
    - (re-)identification, in which people detected in one camera of the camera network are assigned to people detected in another camera (with or without overlapping view), or in which people leaving the field-of-view of a specific camera and then re-appearing in it (e.g. after turning back) are identified to be the same person. This is very important to reduce double counting, analyse trajectories etc. Such (re-)identification is based on global visual characteristics of the detected people (e.g. type of clothing, color, general appearence,...) so no real identity (name of person) is established, i.e.

Question 8. Can you indicate whether there are other data standards (other than Public Life Data Protocol, 2017) that could be more suitable for the Muntstroom project? (Open Question)

Different answers were given to this question.

- Public Life Data Protocol is a good standard to start with. As people use new ways of moving around all the time (emergence of the steps, etc.), models and protocols will therefore have to be adapted or adaptable regularly. Research investigates also environmental and circular metrics in Public Life. They could be added to those developed by Gehl institute and MIT.
- The General Transit Feed Specification (GTFS) is a data specification that allows public transit agencies to publish their transit data in a format that can be consumed by a wide variety of software applications.
- There's a standard called NGSI-LD[7]. People Flow already has a described object-model in this standard.
- Check Flanders Open City Architecture (VLOCA)[8].
- Public Cloud Image recognition protocols. E.g. Amazon and data /metadata retention protocols.
- It seems like there is already really good coverage of all the items tracked in the PLDP in the common schema.org Linked Data schemas[9]. Switching from a custom defined protocol to general Linked Data would be more widespread and known and thus it may be easier for third parties to adopt to. And may help data fusion from different sources that may come in the future.
- The system may detect people who are in panic or patient to realize behaviour analysis functionality

[7] https://fiware-datamodels.readthedocs.io/en/latest/ngsi-ld_howto/index.html

[8] https://www.imeccityofthings.be/en/projecten/vloca

[9] This is mentioned on github: https://github.com/gehl-institute/pldp/issues/3

- The first challenge will be to ‘separate’ pedestrians from non pedestrians, which is already possible today using AI on video. To which extent the full protocol can be measures is to be tested. Other technologies that work on indirect factors (such as speed of movement) will not be accurate enough to capture enough details. The more detailed the analyses required, the smaller the groups will become and the more privacy aspects may pop up.
- There are standard to follow groups of individuals (= crowds) in time. When using this it would also possible to attribute a fragment to objects, allowing people and objects to be "welded" together (shops, public banks, rubbish bins, traffic signs, etc.), which makes it possible to follow a particular individual over time.

Interoperability

The Muntstroom PCP Group wants:

- *that the output of different sensor technologies can be combined and analysed.*
- *that the output of different People Flow-related sensors could also be combined and analysed in all systems.*
- *to receive, understand and (re)utilise sensor-data.*
- *to prevent that during PCP-phase 3 (Living Lab-phase) there are too many sensors placed at one place.*
- *to prevent that sensors only work on one back end.*

To achieve this, the Muntstroom PCP Group intends to promote interoperability.

Question 9. How would you make sensors of different providers work together in an end-to-end solution? (Open Question)

A short summary of the many remarks that were provided:

- Most of the Muntstroom use cases do not require sensors to communicate between themselves. As such, ensuring seamless integration of sensors of different providers deals mostly with how the data output by the sensors is ingested, stored and managed, in a self consistent and coherent manner.
- You can utilize two approaches
    - early fusion in which you merge the sensor data among different sensors from different vendors and sources and then analyse the data together.
    - late fusion
- Sensors have to be connected using one of the few standards available to communicate in smart cities, this is true at each layer of the architecture. A complete standards list must be defined in the project to ensure that sensors are adapted to the communication platforms
- Different sensor providers likely follow different standards when it comes to the data formatting and storage. In order to ensure a seamless end-to-end integration, it is important to achieve standardisation: All data formats should be standardized. This can be achieved, for example, as a preprocessing layer during data ingestion, where data from various sources is standardised before it is stored. Standardisation puts a special emphasis on the use of Open Source formats and technologies as standardising data stored in proprietary formats can be difficult.
- Data streams coming from IoT sensors, real-time data, open data or commercial information can all be ingested in a unified way on a platform. Then a databroker can provide the ability to the data owners to share their data with multiple consumers. So the data ownership is key, but other parties or applications can easily “subscribe” to several data streams. E.g. one sensor or camera can provide information to several applications.

Question 10. How would you measure interoperability in a procurement process? What are the KPIs? (Open Question)

A summary of the different remarks that were provided:

On interoperability:

- As data standardisation is a crucial requirement for the success of this project, one of the key points which should guide the procurement process in terms of interoperability is:
    - Reliance on Open Source formats and technologies.
    - Adherence to industry standards and protocols.
- Make sure no hidden vendor-lock in. No intermediate vendor specific data sharing platform
- Open question to define architecture of sensor providers to integrate into a 3rd party centralized platform.
- Comprehensive API with documentation and support.
- It seems delicate to take a position on this point, as it remains an open and chronic problem in all markets, despite the existence of standards on the subject. Determining KPIs would be tantamount to establishing a hierarchy of degrees of interoperability, on a technical (theoretically objective) basis, but the fact that several standards exist seems to lend credence to the thesis of interoperability "by domain", which is de facto contradictory to "total" interoperability. More an objective of means than results, this ambition seems to us to be difficult to quantify/measure, except by imposing formats ex ante, or by using Open Source by default moreover.
- Contemporary data protocols should be supported and future protocols should be easily implemented in a modular way. Also, on an egress side, standard protocols should be used / supported.
- A Service-oriented Architecture Interface should be provided for external data exchange.
- In the tendering procedure you can ask the different providers to explain the data model and to provide a JSON scheme of the “outbound API”. See e.g. the VLOCA website.
- Create a test bed which either tests the sensor compliance, or the software compliance to the interoperability. These test beds can be made available online.

On the KPI’s

- Is the sensor provider using exclusively proprietary data formats? (binary)
- Is the equipment already following established hardware, data collection and storage standards? (binary)
- Data
    - accuracy of the raw data,
    - accuracy of the extracted data (i.e. data after algorithms are applied),
    - relevance of these data for the result of the work package. That way data with a strong correlation to the result are stronger emphasized and data of lower relevance / meaningless data are marginalized.

- Quality of exchange: The quality of exchange is determined when the exchange is carried out correctly, i.e. when the information sent to the partner is efficient. The quality of use reflects the amount of information obtained by the partner as opposed to the number of the requested information.
- Connectivity: Connectivity can be calculated directly by counting the number of messages received by all the units involved and the number of messages received by the network or by the data link.
- Interoperation time: The interoperation time refers to the period between the date of the request for information and the date of use of the requested information.
- Capacity: A system's capability is the pace at which data can be passed over time. A maximum data rate can be determined for any system or group of systems based on its operating parameters.
- System overload: When more data must be transmitted than the system is able to transmit, a system overload occurs.
- Underutilization: This happens when the data rate/message load of the system is less than its maximum capacity, but messages are waiting for transmission in queues.
- Under capacity: Under capacity occurs when messages linger in queues and the maximum rate of device data is reached.
- Data latency: is the period elapsed by the consumer from the moment of the occurrence to the time of receipt (tactical data processor).
- Interpretation and use of information: The next step will be to ensure that the correct action is taken after the data has been passed and correctly interpreted.
- Conformity: The standard of conformity applies to the use of the data i.e. whether or not the information obtained is exploitable.

Minimal Total Costs of Ownership

The Muntstroom PCP Group would like to control and minimize the Total Costs of Ownership (TCO). As a prerequisite this means that:

1. The sensors need to

a) be as stable as possible
b) be easy to install and remove, maintained and/or replaced
c) have a minimal susceptibility to interference
d) have a long lifespan
e) have a CAPEX and OPEX that are as low as possible.
f) that have limited telecommunications and interoperability costs

2. The intelligence and innovation need to be, in particular, in the platform and not in the sensors.

Question 11. Which general recommendation(s) could you give on the relation between the Total Cost of Ownership (TCO) and sensors? (Open Question)

A summary of the remarks that were provided:

- The costs also depend on the regional business case. If, in the large-scale deployment phase, the number of sensors go beyond a certain number the license cost of the software starts decreasing significantly.
- Sensor takes advance rapidly, and concur with the idea of having intelligence at the platform. Yes, some sensors could come with edge capabilities, good. Nevertheless, pushing too much intelligence to edge is not conducive to an end to end, shared process.
- The most valuable raw data for all use cases is live video. However, you want probably to achieve a minimum data communication? The solutions regarding the proposed use-cases should then provide sensors with the maximum capability for on-edge processing and using smart ways to store critical data. This way the TCO will be held as low as possible in terms of purchasing and maintaining sensors wherever camera-based solutions are crucial.
- Take the practical issues of sensors in mind:
    - Saving battery (not using rechargeable batteries) is saving replacement period. This is the reason why modifying sensor parameters should ideally be done remotely.
    - A conscious choice can be made for sensors that do not require cabling.
    - Consideration should be given to the influence of the positioning of the sensors and their encapsulation in order to minimise vandalism.
- Sending all sensor data to the cloud and processed there, requires the right compression to the data and a protocol that is as light as possible. However, all together, it could still lead to inefficiency in performance (due to latency) as well as cost price (data communication and storage)

- A platform should be able to handle two extreme cases:
    - Pushing real-time streams (data or video) towards the platform and applications. These applications can be data analytics tools, visualisation tools, … all with their own processing methods and techniques.
    - The processing of data on the edge and only send the events, alarms, triggers, metadata to the platform. Not only to do for example: object recognition, but also it can be important to apply some hashing techniques to make sure that privacy sensitive data isn't processed, kept, … on the platform.
- The "smarter" the sensor, the higher the TCO basically.
    - The physical installation is quite susceptible to malfunction due to many factors (weather conditions, vandalism, power issues, etc …). Preventing these can have an impact on the TCO.
    - There is also a big difference between different sensors: The reach of a sensor, the number of sensors needed, the amount of processing needed etc … can also have an impact.
    - In order to keep the TCO under control, you could make a smart combination of sensors and technologies. Does the level of detail have to be available in a complete area, or is it enough to have the highest level of detail in specific places only?
- Finding the best price-quality for sensors is not easy.
    - It is not useful to buy the most performant sensors because technology evolves rapidly.
    - Choose sensors that require minimal maintenance. For instance, explore the possibility to use self-powered sensors.
    - Choose the cheapest IoT Protocol & Operator depending on the frequency required for sending data
    - The choice should also be linked to the data quality and details that should be measured according to the use cases.
    - Conclusion: different sensors have advantages and disadvantages. Make a list of what each sensor can do and cannot do and for which use cases it is well suited.
- By using a platform based solution, you can unleash the power of a supercomputer which TCO wise is impossible in the edge. This also allows to reduce the cost of sensors as these can be reduced in ICT intelligence and footprint.
- Difficult to answer the question as long as the size of the living lab is not yet known... How many sensors? Which geography? Which type? In the current project phase, the sensors should not exceed 15% of the TCO.
- This architecture seems to suggest all sensor data is send to the cloud and processed there. This could lead to inefficiency in performance (due to latency) as well as cost price (data communication and storage)
- Consider stripping out the physical sensor installation to existing infrastructure partners as to leverage existing structures and processes and to operationalise the scaled solution more efficiently.

4.3.2 Communication layer

Open standards

Question 12. The Muntstroom PCP Group intends to promote interoperability. To achieve this, it wants to mandate open communication standards. How would you arrange this? (Open question)

A summary of the remarks that were provided:

- Open communication standards are important to achieve that a platform is interoperable and can be integrated with other third-party components and applications. Therefore, a platform should support flexible, scalable and open architecture. With as key principles: Open Architecture, Modular Design, and Common Data Object Approach.
- The idea would be to select and use open and commonly-used communication protocols and API. An example is the REST API over HTTP protocol for communication over the web.
- This can be achieved by constructing a standardization layer in the data infrastructure before the data storage. This layer would support conversion of data from a variety of sensor types into standardized forms, and ensure that the data which is stored and available to the end users is uniform.
- A layer of metadata using a common 'ontology' can stimulate interoperability. While such standards exist (e.g. Flemish OSLO standard, BIM standards …).
- Learn from other projects (TMaaS, Mobilidata,...)
- In the design phase discuss which data format will be used (make a cost/benefit analysis)
- Build a strong and open development community around the tools and try to actively maintain and support it.
- I'm not sure if communication standards are relevant to interoperability, I think it's more open end points and standards rather than communication standards, especially correlated to cost
- First level is to work on open network standard, for example LoRa should be quite nice choice for area wide coverage. On top, open network protocols and APIs should be accessible for general use.

4.3.3 Storage layer

Storing different sorts of data

The desired solution needs to be flexible and secure to store different sorts of data like:

a) personal and non-personal data
b) data from different sources (via B2G or G2G[10] data sharing)

Question 13. How would you comply with this flexibility and security in the system? What are the challenges? (Open Question)

A summary of the remarks that were provided:

- By complying with security on three levels: Operating system, Codes and Data transmission.
- Our proposed solution has the following features:
    - Scalable: could be extended in case of low capacity.
    - Fast: Minimum latency to access data.
    - Reliable: Prevent data loss and resistant to potential hazards like bad sector.
    - Hierarchical Data Access: Let anyone access their data.
    - Minimum Redundancy: Prevent duplicate data.
- The data at rest should be kept on European soil or strict conditions (safe harbour principle) for data storage should be followed.
- By implementing data ownership and enforce GDPR compliancy. Except for open data, all data on a platform will have a data owner. It's up to the data owner to share his data in an open or commercial way with one or many subscribers. Concerning personal data, there a several techniques and tools that can be used by data owners to protect their data.
- Complying with cyber security and information security standards such as IEC62443 and ISO/IEC27001.
- One way to ensure that the personal data is protected is to allow only access to aggregated and anonymised personal data. Alternatively, the data infrastructure could serve already anonymized and aggregated data in case this doesn't have to happen at the time of data request.
- Ethical data uses: Ensuring that the data obtained by Muntstroom is used in an ethical way is another challenge. Considerations of questions about ethics and fairness should be an integral part of the data storage and end solution development process. This can be done by requiring an ethics evaluation of all end solutions as a part of the development process, for example via a data/AI ethics framework.
- Compliance can be met by storing the data in way that allows easy distinction of the required user access levels. By creating user entities with different access to the stored data the desired level of flexibility and security can be achieved. Of course, the sheer number of data collected can make retrieval in timely manner challenging.
- The solution may be developed on hash-based encryption technologies. High level encryption methods may be used on data layer to store extracted meta-data.

[10] Business to Government and Government to Government

- Follow the CIA model (Confidentiality, Integrity, Availability) of ISO27001. This implies not only technical competences but also security processes (governance) and awareness. At the technical level a best of breed approach is followed using FIPS/NIST/CIS recommendations
- Deploy a Quality management system/broker that separates/filters data according to regulations. To deliver this with sufficient speed use broker farm.
- Make a split between personal data and non personal data. Handle only the personal data with strongest GDPR and security constraints. This keeps performance high.
- Our suggestion is to not store personal data at all. Main principle here is that only data that are not stored are not able to be misused. On top, end-to-end encryption from sensors to end user should take place.

Question 14. What KPIs would you use to measure these requirements? (Open Question)

A summary of the remarks that were provided:

Security KPI(s)

- security ratings: Security ratings are often the easiest way to communicate metrics to non-technical colleagues through an easy-to-understand score. A simple A-F letter grade based on various criteria including network security, phishing risk, DNSSEC, social engineering risk, DMARC, risk of man-in-the-middle attacks, data leaks, and vulnerabilities.
- Patching cadence: How long does it take your team to implement security patches or mitigate high-risk CVE-listed vulnerabilities?
- Access management: How many users have administrative access?
- Non-human traffic (NHT): Is there a normal amount of traffic on the server or is there an uptick that indicates a potential body attack?
- Cost per incident: How much does it cost to respond to and resolve an attack? How much money are you spending on staff overtime, investigation costs, employee productivity loss, and communication with customers?
- Mean Time to Detect (MTTD): How long do security threats fly under the radar at your organization? MTTD measures how long it takes for your team to become aware of a potential security incident.
- Mean Time to Acknowledge (MTTA): What is the average time it takes you to begin working on an issue after receiving an alert?
- Mean Time to Contain (MTTC): How long does it take to contain identified attack vectors?
- Mean Time to Resolve (MTTR): How long does it take your team to respond to a threat once your team is aware of it?
- Mean Time to Recovery (MTTR): How long does it take your organization to recover from a product or system failure?

Flexibility and Scalability KPI(S)

- Administrative scalability: The ability for an increasing number of organizations or users to access a system.
- Functional scalability: The ability to enhance the system by adding new functionality without disrupting existing activities.
- Geographic scalability: The ability to maintain effectiveness during expansion from a local area to a larger region.
- Load scalability: The ability to expand and contract to accommodate heavier or lighter loads, including the ease with which a system or component can be modified, added, or removed, to accommodate changing loads.
- Average Scale-up Time: Average time it takes to add a new node and the new node is ready to work.
- Number of requests handled by a single node.
- Maximum number of nodes.
- Maximum concurrent users.
- Throughput: Minimum latency to access data using a single node.
- Generation scalability: The ability of a system to scale by adopting new generations of components.

Reliability KPI (s):

- Mean Time Between Failure (MTBF): average length of operating time between failures for a specific piece of equipment or component.
- Maintenance cost per unit (MPU): total maintenance cost divided by the number of produced units in the measurement period.

4.3.4 Processing layer

For the combination of data and the basic processing (like a quality check) software needs to be developed. The Muntstroom PCP Group wants this software to be Open Source (OSS), so other contracting authorities can use, reuse and modify the software, in accordance with their needs.

Question 15. Do open standards (Open Source Software) for the combining and basic processing of sensor-data already exists, (yes/no)

	Answers
Yes	23
No	6
No answer	10

If yes, can you give some examples?

- CityFlows[11], OpenRemote, Tensorflow, Python, Pytorch, OpenCV, etc.
- First there is a difference between open standards (which promote interoperability and avoid vendor lock-in) and Open Source (which is one type of commercial approach to software/hardware procurement). The two are mutually exclusive. Fiware and OGC provide some examples of open standards. These standards are implemented both as Open Source and open software components.
- Most sensor data is collected in "classical" format: SQL, CSV files, flat files, ASCII, syslog, Radius, . This mainly depends on the vendor choice and architecture. The more usual one is typically sensor data from (Video Management). VMS systems which need to be addressed through SDK. That said, major VMS vendors have been natively integrated in the above processing layer
- Another important topic (before the data) is the platform itself. This should be an open model / as open as possible for Analytic layer players. At first this could be virtualisation (VMWare, Hyper-V, ...) but given the way forward you could move to the next level of platformication, i.e. containers (Docker/Kubernetes)
- Below the virtualisation/container layer the IaaS too can be transparent being either customer sited (HP/Dell), Private Cloud or Public Cloud (Azure/AWS).

Question 16. What KPIs would you use to measure these requirements? (Open Question)

A summary of the remarks that were provided:

- Specific criteria for open standards have been defined by the Open Source Initiative (OSI).
- Open standards is binary: certified or not
- Data Integrity: there is no loss of data and it remains relevant.
- Data Consistency: the data is uniformly coherent.
- Data Completeness: there is no missing data.
- The number of "translations" necessary (in Dutch: vertaalslagen).
- The required processing time, scaling capabilities.
- the balanced score card methodology.
- Accuracy, Improvement
- The implementation of the processing should be based on microservices to enable stability. Also it is recommended to rely on a distributed processing engine like Storm.

[11] https://www.imeccityofthings.be/en/projecten/cityflows

4.3.5 Analytics layer

The required analytics are defined by the use cases. Each use case presents a different scenario for analytics, with possibly different data sources. See the use case-document and the questions in the following block.

Each use case presents a different scenario for analytics, with possibly different data sources.

Question 17. What consequences / challenges would this raise for e.g. the architecture? (Open Question)

A summary of the remarks that were provided:

- The type of data have to come from various sources, various type and at various speed. The level of data 'choices' at get go should be clearly defined. hence, what data to use is the critical questions. The second difficult part would be to answer with the right analytics tool.
- You need to provide agility in ingesting any data source while data interpretation and processing is done within the cloud platform.
- In some cases, being able to react immediately to new data is more important than being 100 percent certain of the data's validity. The architecture is a data processing design pattern to handle massive quantities of data and integrate batch and real-time processing within a single framework.
- Compositionality of AI models (DUET)[12]
- The main challenge is that the Analytics layer needs to have a modular design in order to handle the various proposed used-cased. It will use different algorithmic techniques (clustering, tracking, graph prediction, etc.) that can be used and shared by the different use-cases.
- When several uses cases wil depend on different and similar data sources an appropriate approach is to create a centralized data storage, with structured data and non-structured data (data lake), and avoid data silos for each use case. Data silos have multiple drawbacks such as data replication, data inconsistencies, expensive cost, etc.
- I think the number of use cases is too high, and that the architecture could become too complex to try to support it all
- With a modular, micro-services architecture you ensure that challenges can be highly localised, and their solution can be specific tools.

[12] https://www.digitalurbantwins.com/flanders-twin

4.3.6 Smart Access layer

The Muntstroom PCP Group would like the desired solution to support different business models. When you combine this requirement with the earlier requirements on the storage and analytics, a challenge emerges. The public buyers have a couple of questions about this.

Question 18. How can you ensure that the output data per use case complies with the sharing conditions of both public and private data sources? (Open question)

A summary of the remarks that were provided:

- By making all data, besides open data, owned and under the responsibility of the data owner. By enforcing the use of data processing registers and processing agreements. And by ensuring that the data owner can share it's data in an open data portal, commercial portal or privately, in any presentation or format.
- Check the GAIA-X initiative, which answers these specific questions[13].
- With a Strong Authorization system, you can ensure that all data would be safe under both conditions. Public data will change due to your needs, for example, anonymized and private data will be accessible by selected users inside the government.
- A suitable method may be to restrict the public layer of the data. In this way, the contents of the public layer can be specified and public access can only be provided to the specific items. For example, access to data about the number of people crossing a street (any street) can be given to the public. The public will, however, only have access to the number of people crossing a street and not to more details.
- At a lower level in the architecture, personal data must be shielded from this layer. This can be done by providing a data layer that only contains anonymised data.
- The use of data auditing pipelines and a framework in which privacy-sensitive data is identified and simply not presented towards systems that end up publishing the final public data layer. Furthermore privacy-preserving techniques such as Synthetic data generation could be applied to generate sets that have the some statistical properties but ensure privacy-preservation
- By guaranteeing both open and private access. Governance between the two can be assured through the implementation of an ISO27001 governance process. Data ownership will float so DPA is required. It is relatively simple to implement DPA between the data owners and one company. But a lot more complex towards the "data consumers" i.e. government, public, personal. The complex part is not a technical problem but more the privacy/compliance aspect.
- It might be the case that data from model to model are not different in terms of content but in terms of timeliness (real-data delayed data, frequent data versus lower frequent data. This is why also different stream patterns for each business model might be applied.
- The most pressing challenge is to select the right IoT technology stack for the enterprise with the scalability to meet your current and future needs. There is however a lack of IoT open standards and interoperability across platform vendors. So you need a flexible and modular architecture to quickly respond to changing market conditions.

[13] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

Question 19. Can you suggest measurable values / KPIs for the different business models (Closed, Shared and Open from Figure 3)? (Open Question)

A summary of the remarks that were provided:

- Check the GAIA-X initiative, which answers these specific questions[14].

Some KPI's

- Number of users.
- Number of commercialized projects.
- Number of added value projects.
- Number of governmental contracts.
- Number of commercial contracts.
- Number of new aggregated Datasets created.
- Number of policy changes-improvements based on the obtained data.
- Accuracy of data over 95%.
- Number of research publications based on the obtained data.

4.3.7 End-users layer

Question 20. How can you ensure the different level of access to different end-users, while being compliant with Data Protection Regulation. (Open question)

Different answers were given to this question:

- A proper authentication and authorization system should be setup and be continuously monitored. Probably additional techniques such as user impersonation may be used while setting up roles and assigning users to them in order to evaluate the results of each administrative action.
- Clear separation of data and data access models (role and rights management in combination of physical data separation).
- By supporting different user categories and ensuring that any data presented on public interfaces only presents aggregated trends and not individual user's patterns, not even anonymously. Because this could be used maliciously to track individuals.
- Present a certification in the trend of ISO27001 or similar. Be aware that all employees of data handler and data owner are regular trained on GDPR and that training records are kept with integrity.
- By creating a data mapping and flow document in the initial design phase and keep it current throughout the deployment and lifecycle of the system.
- At a lower level in the architecture, personal data must be shielded from this layer. This can be done by providing a data layer that only contains anonymised data.
- Each access point type should ensure to extract, aggregate and potentially anonymise only the data appropriate for a particular user type.
- Terms and conditions explain why and how one uses which data and how data is collected.

[14] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

- Private data should be encrypted, at minimum at rest but better would be in transit and compute as well. This however (especially compute) has serious implications on the development of the framework.
- Identity and Access Management (IAM) mechanism.
- Hash technologies.
- Aggregated data are not considered personal data. Regarding raw data, the images, using thermal cameras means that even those images may be shared with any end-user.
- Role-based access control (RBAC).
- By applying Dynamic Masking (DDM). Technology that aims at real-time data masking of production data. Dynamic data masking helps prevent unauthorized access to sensitive data by enabling customers to specify how much sensitive data to reveal with minimal impact on the application layer.
- Small data approach for personal data is the right one, but it has to be approached in a decentralised way through linked data organisation and ready-to-be inferred normalised data models, no matter the data source. Applying a governance framework to ensure the end-users' control of the data all along the lifecycle is mandatory. Processing of the information must be done in confidential computing environments allowing to extract the necessary insights to nurture the application, this approach allows the service provider to remain fully compliant. Accountability and transparency must be stored in open ledgers and anytime revoke access must be provided to the end-users.
- By discussing
    - whether it is necessary to work with personal data. Nothing can ever go wrong with what you don't have.
    - how to keep the personal data as short as possible.
    - how to only give access to those who need to be able to work with it.
    - how to guarantee that data is deleted it.
    - how to return data unaltered if you ask.
    - what to do with data if it needs to be kept for a longer period of time.
    - how to avoid theft.
    - etc.
- The log is one of the most important elements of the security infrastructure. In this way you can check for everything: how long you have had it, who worked with it, what the original information was, when and how often it changed, and so on. So it offers you full control and transparency.

Question 21. Which requirements can you suggest for the end-user layer?

Different answers were given to this question:

- For the end-user layer, you need a feature-rich panel that supports most of the APIs. Everyone at any level of access to the data should get their data in every format they want. Formats like JSON, XML, etc would be appropriate.
- Role and rights management.
- A Data Protection Impact Assessment (DPIA) should be applicable to the end-user privacy agreement which will ensure transparency and independence.
- Should have different (adjustable) user categories, access policies and groups.
- Besides a system under real-time processing, it should also provide flexible frontends which can work on web, cloud and mobile platforms.
- Identity and Access Management (IAM) mechanism linking profiles to user rights to underlying systems and data.
- Use of e-id for Belgian end-users.
- Working with personal data stores (user centric data lakes opposite to classic organisations' centric ones) and consent templates allowing to also normalise the legal framework along with the data itself.
- The Open Web Application Security Project (OWASP)

4.4 Use cases (required analytics)

The required analytics are defined by the use cases.

4.4.1 Use case approach

Question 22. Is the approach of use cases in the context of PCP clear to you? If not, what would you have explained (differently)? (Open Question)

Most respondents responded with "yes" or left the question open. Some specific remarks:

- In general, the approach is explanatory and relatively easy to assess. What we would like to be more thorough explained is the requested technology used for monitoring. For example, are data gathered by cameras included?
- At a high level, this is clear. The concrete implementation is a bigger issue.
- Yes, but there is a risk that the amount of processing required is far too great to cover all the needs at a reasonable cost.

4.4.2 Remarks on the use cases

Question 23. Do you have a remark on any or several of the use cases?

Pedestrian analytics

- Very important use case as it is a basic building block for many other use cases.
- Various respondents say they are already familiar with the use case.
- Why only pedestrians? Better multimodal analytics. Additionally, it is hard to assess if a pedestrian is running or walking.
- The granularity of categorizing different kinds of pedestrians (walking cane vs. guide cane, strolling vs walking at average pace, etc) will be challenging.
- There is a link between the very granular detection vs. the return on investment.
- Complication: training AI models (image CNNs & video CNNs) requires a lot of annotated data. Both the collection of sufficient data and the annotation of this data is a very big effort. Certain individual features (e.g. 'gender', 'age') and group dynamics ('groups', 'activities') are hard to asses. Successful classification assumes the existence/development of a lot of building blocks such as object tracking (to gather person-centered video frames), object mapping (image coordinates-to-world-coordinates) to asses speed etc.
- The first challenge will be to 'separate' pedestrians from non pedestrians, which is already possible today using AI on video. To which extent the full protocol can be measured is to be tested. Other technologies that work on indirect factors (such as speed of movement) will probably not be accurate enough to capture enough details.
- The more detailed the analysis required, the smaller the groups will become and the more privacy aspects may pop up.
- Interesting part for this use case can be to give an indication of how crowded a place is before someone is planning to visit it.
- Identifying tourists versus business travellers could be useful to certain extent (i.e. groups identification).

Analysis of reasons of movements

- This requires access to multimodal data (ticket data, timetables, wifi users, etc) and therefore close collaboration with the public transport operator is needed.
- This kind of use case requires a longitudinal view (same ID used over time). This can be offered by Wifi / mobile location analytics. Then profiling can be done based upon visited places, stay time, frequency, most likely living places etc … These pose challenges on privacy & GDPR compliance, since reporting can only be done on aggregated groups of profiles. This can be enhanced with spot measurements based upon camera imaging analysis for more detail. Basic requirement: possible; Bonus requirement: R&D needed.
- Due to the correlation with event data (plans, arrivals, departures) and a longitudinal view (same ID used over time), there is a potential risk of not having enough data sets to make this assessment.
- The reason is not measurable, if somebody runs, can be multiple reasons; hurry to catch bus, afraid of dog or simply jogging. Easier to figure out if someone is visiting a grocery shop but hard to find out if someone goes into the subway.
- Basic requirement #2 seems to be impossible, e.g. difference between workers, passers-by and local inhabitants. Whatever the algorithm, answers will only be possible with a certain confidence.

Events analytics

- This can be helpful for organizers to understand how to make better decisions for the event location and other arrangements like food, chair, table, etc.
- The use case needs correlation with event data (plans, arrivals, departures) and multimodal and real time of events.
- It might be interesting to look at Digital Twins, for simulation purposes ("what will happen to the people flow if one places this object there").
- Event analytics should be approached in a layered approach. Each layer can generate different insights.
    - On macro level geographically granular information (e.g. from Wifi, sensors, …) can be combined with mobile location analytics.
    - Short distance through Wifi technologies and using sensors.
- Bonus requirements seem very complicated. Unclear what is meant by '(un)desirable'.

Crowd analytics: security

- Seems feasible on an aggregated ("crowd") level, but it seems less desirable to zoom in on individual behaviour ("person").
- This can already largely be done with either existing data from carrier networks or with wireless sensors and RF signals to measure crowd density.
- The bonus requirements are very complicated (e.g. predictive analytics). To train the predictive models you need years of historical data.
- Map and organisation dependent thresholds.

Crowd analytics: Covid

- This must be assessed in term of process on top of the technology. What can you do with this information? Which information is relevant so notifications do not become annoying?
- Seems feasible on an aggregated ("crowd") level, but it seems less desirable to zoom in on individual behaviour ("person"). So this use case is very feasible if you want to limit the amount of people in a certain zone, not feasible if you effectively want to make sure the 1.5m distance is kept. This will result in a lot of false positives, as you cannot train an algorithm to know which people belong in the same bubble and which don't.
- To increase the accuracy of the output, additional mobility data that may be available (e.g. taxi and electric scooters fleets data, social media geotagging metadata, aggregated telecom data, data from bluetooth traffic sensors) could be used.
- Indoor and outdoor routing services for pedestrians which will suggest the safest route taking into account the zones of increased concentration of citizens may be also be developed.
- Density and distancing.

Mobility hub analysis: usage of specific PT-lines

- This requires correlation with multimodal data (ticket data, timetables, pubic wifi users, commuter-specific touchpoints, mobile app subscribers, etc) and therefore close collaboration with the public transport operator is needed.
- De Lijn has already worked on a PoC for counting alighting and onboarding passengers on vehicles with small cameras above the doors of the coastal trams.
- Before starting I would suggest to check which data is already available and well define how accurate the analysis has to be.
- Can sensors be placed in the vehicles themselves, or should the analysis happen from fixed sensors at the stop places?
- Also finding pedestrians with and without an abonnement can be interesting to figure out with mobility hub analytics.

Mobility hub analytics: transfers

- Finding everyday distance travelled between stations using public transport can be helpful in calculating the carbon footprint with comparison to private transport.
- More complex than use case 6, since this involves people tracking + “nature of conflict” between pedestrians.
- This requires access to multimodal data (ticket data, timetables, pubic wifi users, commuter-specific touchpoints, mobile app subscribers, etc) and therefore close collaboration with the public transport operator is needed.
- For this use case, there can be a layered approach:
    - Long distance trajectories using mobile location analytics.
    - Short distance through Wifi technologies.
    - Spot measurements using sensors / sensors.

Asset management

- This is an innovative use case, as the behaviour of people can change over time and thus the need for prescriptive analytics.
- When looking at the use of the assets within the mobility hub, cooperation with the transport operator is needed.

Commercial analytics: shopping policy and research

- This is a use case with which various suppliers say they are already familiared with.
- Finding location with the most visits and time spent can be included in searching commercial potential locations.
- Some of these requirements require data that is, according to us, beyond your control. So this use case is probably more than “average” in complexity.

Commercial analytics: individual shop

- This is a use case with which various suppliers say they are already familiared with.
- Idem to the mobility hub use cases: if you want to build heatmaps of in-store movements, you need collaboration of stores and permission to use their data.
- Nothing is mentioned about augmented reality on the smartphones of pedestrians or promo messages on the smartphones. Also, there is no link with the online / promo campaigns of the concerned POS which are maybe more important than the vitrine message of the POS. Attractive outside view with boards and different discounts display can be potential pedestrian stoppers.

Traffic light analytics

- This is a use case with which various suppliers say they are already familiared with.
- The output of this analysis may also be extremely useful to the vehicles approaching those specific traffic lights (i.e. to enhance pedestrians´ security by alarming the driver under certain circumstances). For this communication, additional interfaces development may required.
- High complexity due to basic requirement 'nature of conflicts' between pedestrians.
- Unclear: operational definition of a bottleneck.

PMR routing: Wheelchair

- This is a use case with which various suppliers say they are already familiared with.
- This is a more difficult use case because you need to have sufficient examples to analyse properly. Are there already step free routes available from the metrostation Debrouckère, as a part of the direction plan, which can help wheelchair persons to travel easily?

PMR routing: Partially sighted person

- This is a more difficult use case because you need to have sufficient examples to analyse properly. Are there already step free routes available from the metrostation Debrouckère, as a part of the direction plan, which can help handicapped persons to travel easily?
- Voice enabled features within an application make routing easier for partially sighted.
- Privacy issues are strong here.

Indoor routing: Subsurface

- This is a use case with which various suppliers say they are already familiared with.
- Analyzing indoor routing is very different from analyzing city routing, it requires different solutions and data sources.
- Cases 14 and 15 are the same...a subsurface construction is similar to a building with many floors.
- Subsurface monitoring could well be too challenging for this case.

Indoor routing: Building

- This is a basic use case with which various suppliers are already familiared with.
- Analyzing indoor routing is very different from analyzing city routing, it requires different solutions and data sources.
- Problem is when the local network is shortly interrupted. Experimenting with indoor routing technologies based upon various technologies could be a solution.
- Maybe foresee a building -level sensor aggregator role.

Smart Access

A summary of the different remarks that were provided:

- You could make a link to the GAIA-X initiative[15]. It answers these specific questions.

[15] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

4.4.3 Complexity and testing

Question 24. Could you indicate the complexity of the use cases by ordering them as basic (1), intermediate (2), advanced (3).

Use case	1	2	3	No answer
1. Pedestrian analytics	10	14	9	6
2. Analysis of the reasons for movements	3	8	21	7
3. Event analytics	5	22	6	6
4. Crowd analytics: security	5	17	11	6
5. Crowd analytics: COVID	10	18	5	6
6. Mobility hub analytics: Usage of specific public transport lines	5	19	9	6
7. Mobility hub analytics: Transfers via gates, platforms and corridors	5	17	10	7
8. Asset management	8	16	9	6
9. Commercial analytics: shopping policy and research	5	16	9	9
10. Commercial analytics: individual shop	8	13	10	8
11. Traffic lights analytics	9	16	6	6
12.Routing PMR: wheelchair	6	13	11	9
13. Routing PMR: partially sighted	2	14	14	9
14. Indoor routing: subsurface	4	14	13	8
15. Indoor routing: building	9	12	9	9
16. Smart Access	6	11	13	9

Table 1: Order by number use case

Use case	1	2	3	No answer
2. Analysis of the reasons for movements	3	8	21	7
13. Routing PMR: partially sighted	2	14	14	9
14. Indoor routing: subsurface	4	14	13	8
16. Smart Access	6	11	13	9
4. Crowd analytics: security	5	17	11	6
12.Routing PMR: wheelchair	6	13	11	9
7. Mobility hub analytics: Transfers via gates, platforms and corridors	5	17	10	7
10. Commercial analytics: individual shop	8	13	10	8
6. Mobility hub analytics: Usage of specific public transport lines	5	19	9	6
8. Asset management	8	16	9	6
9. Commercial analytics: shopping policy and research	5	16	9	9
1. Pedestrian analytics	10	14	9	6
15. Indoor routing: building	9	12	9	9
3. Event analytics	5	22	6	6
11. Traffic lights analytics	9	16	6	6
5. Crowd analytics: COVID	10	18	5	6

Table 2: Answers ordered by complexity

In PCP-phase 3 the developed prototype will be tested in a real environment. The foreseen test period is half a year and the solution will be tested by means of some of the use cases.

Question 25. Which use cases would you focus on? Why?

A summary of the different remarks that were provided:

- For many respondents: Use case 1. Pedestrian analytics seems to be the basic use case.
- Additional use cases are chosen in relation to:
  - The knowledge and experiences of the respondents (from both analytics and technology perspective).
  - The (too) high amount of annotated historical data needed to train predictive/prescriptive analysis models.
- Different respondents suggest as phased approach:
  - Start with the simple analytics use cases, then add complexity.
  - Start with the routing use cases: these use cases don't require artificial intelligence (routing is optimization) and are less sensitive to GDPR and personal data.
  - Let's start with 1. pedestrian analytics as in terms of technology. With this use case you can test the deployment of the basic technologies needed for all other use cases.
- All use cases are complementary but those that enable behavioural insights are more valuable.
- Some respondent suggest to limit the amount use cases (e.g. one basic and two additional use cases) because:
  - Many tasks have a high complexity and the foreseen R&D budget (500 KEUR) is limited. Dispersing this over too many use cases will put an unreasonably high contribution demand on R&D partners.
  - Obtaining such a great size of date is a of little relevance if they are not effectively and efficiently analyzed, allowing the export of valuable information and insights.

Question 26. What methodology would you recommend to test the use cases?

A summary of the different remarks that were provided:

- You could ask for comparison dashboards to show results of the use cases.
- Prepare ground truth for each of the cases and testbench the output for each product on these cases. It needs to capture some raw camera feeds for different scenarios and get annotated them by humans accurately. Therefore, the output of analysis for each case can be verified and evaluated accordingly. Also, the processing speed can be benchmarked with different hardware configurations for performance assessment concerning a live and real-time processing time.
- An agile methodology to allow an iterative process and to rapidly deliver a prototype and discuss the expectations of the client and his feedback.
- Our approach includes creating a test plan and test cases as per the requirements, getting them approved by the stakeholders. The actual testing of the use cases covers the following types of testing:
  - Functional Testing for all the functions and features as per the requirements document.
  - Usability Testing for all end user applications.

- Performance and Load Testing for the end user applications and APIs.
- Security Testing for all externally available interfaces.
- Final User Acceptance Testing before releasing the application to the entire user base.
- Regression testing after every patch/enhancement.

All the requests and defects are prioritized by the project management team with the inputs from stakeholders. A test summary report is provided along with Release Notes after every major release.

- In order to maximise the value generation of the proof of concept phase, it could be beneficial to follow the rapid deployment scheme, where the emphasis is put on deploying a solution to the use case in the minimal amount of time, without spending much resources on optimising the solution (experience has shown that when it comes to IT solutions, the value generated in the first deployment is usually significantly larger than the added value from each subsequent improvement).
- The entire process is best achieved using agile development methods with short iterations (2-4 weeks).
- As no standard solution exists for the Muntstroom use cases, it is unlikely that a standardized test procedure exists either (including test data). Performance of use cases can be tested by periodic comparisons with truth data or by maintaining a database of test data samples which are properly labelled. These can be collected by the sensor equipment themselves and labelled by hand, although this creates an additional need for human resources.
- "Think Big, Start Small, Fail Fast". Adopt an iterative methodology that allows you to launch small experiments fast, which have a clear added value in the overall project goal. Be satisfied with small wins and partly gains, build on them, but don't dwell and what doesn't work.

4.5 Privacy & security

With the desired solution the Muntstroom PCP group would like to "follow the river, but not the little droplets". This means that the desired Muntstroom solution should be designed in line with the current Data Protection Regulation and that an end-to-end solution for privacy and security is needed. See figure 2: architecture

4.5.1 Privacy-by-design strategies

The EU Agency for Cybersecurity (ENISA) has defined different strategies to integrate privacy and data protection principles directly in the design phase. When these strategies are mapped against the data architecture it shows clearly what principles should be in place in the different stages. See the table below.

	Big data value chain	Key privacy by design strategies	Implementation
5	All phases	ENFORCE / DEMONSTRATE	Automated policy definition, enforcement, accountability and compliance tools.
4	Data use	AGGREGATE	Anonymisation techniques. Data quality, data provenance.
3	Data storage	HIDE	Encryption of data at rest. Authentication and access control mechanisms. Other measures for secure data storage.
		SEPARATE	Distributed/ de-centralised storage and analytics
2	Data analysis & data curation	AGGREGATE	Anonymization techniques (k-anonymity family, differential privacy).
		HIDE	Searchable encryption, privacy preserving computations.
1	Data acquisition / collection	MINIMIZE	Define what data are needed before collection, select before collect (reduce data fields, define relevant controls, delete unwanted information, etc), Privacy impact Assessments.
		AGGREGATE	Local anonymization (at source).
		HIDE	Privacy enhancing end-user tools, e.g. anti-tracking tools, encryption tools, identity masking tools, secure file sharing, etc.
		INFORM	Provide appropriate notice to individuals – Transparency mechanism.
		CONTROL	Appropriate mechanisms for expressing consent. Opt-out mechanisms. Mechanisms for expressing privacy preferences, sticky policies, personal data stores.

Table 3: Privacy by design strategies. Source: Privacy by design in big data (ENISA, 2015)

Question 27. Which privacy-by-design strategies do you propose per data-architecture layer?

Please upload a picture of a document describing this

- Anonymize the privacy of people in the sensor stage (which is the earliest stage of the data architecture), so in all subsequent stages no data about the identity of people exists and thus our strategy is in-line with the privacy-by-design approach and it guarantees complete consistency with GDPR regulations.
- Radars do not capture data that can be used to identify people, it is anonymous by its nature.
- Use the ISO27001 methodology to define and document the privacy by design and default.
- An example:
    - Physical sensor layer: First is to acquire data only relevant data from sensors. Second is to encrypt data when it can be. Third is to make people aware of the recording.
    - Communication layer: All data from different sources are collected and stay anonymous. Encryption mechanism make sure that the system is hacking-proof.
    - Storage layer: The access to the data structures is protected by strong authentication mechanism. Again, the data stored are anonymous.
    - Processing layer, application a smart access layers: Every algorithm runs in a protected environment (either in private clouds or in private servers). Statistical analysis are conducted without compromising the data privacy.
    - End-user layer: The access to the platform is guaranteed to each type of end-user by a dedicated authentication mechanism. Techniques like two factors authentication, and id-verification might be used depending on the use-case.

The use cases present a multitude of challenges and target groups where many variables have to be taken into account.

Question 28. How would you apply the data-by-design strategies to the use cases? Can you point out any main differences among the use cases? (Open Question)

A summary of the different remarks that were provided:

- By anonymizing the private data in sensor-level (the earliest stage) there would be no private data in the subsequent stages. This way a solution will be generally “private-by-design” for all the use cases.
- Please refer to the answers about the data layer in section 4.
- All data collection must be done with user privacy in mind. It must be clear of what is collected as well as the fact that there will be no personal profiling. All data will be used for creating a model to help all pedestrians in the city. There will also be also the ability to opt-out of data collection. Of course, this will also be applied to all stages. Also, different data access levels will ensure that each entity will have access to specific or anonymized data collections. This will be the only differentiating factor between the proposed used cases.
- Some use cases don’t need any personal data (all the routing uses cases). All the use cases requiring video processing and detection require anonymisation methods to be compliant to GDPR rules.

- A basic requirement in the "smart access" use case is to let each full user have full access to its own raw sensor data. That implies that the raw data need to be stored in the system even after they are processed? As the majority of the other use cases may not provide direct access to raw data this differentiates this use case between others.
- Some use cases are based on object detection whereas some others based on behaviour analysis. Our design strategy should consider used technologies and models for each use cases.
- See attachment. A structured approach is followed.
- Capture anonymous data as much as possible

4.5.2 Privacy embedded in an end-to-end solution

The Muntstroom PCP Group would like an end-to-end solution that embeds privacy & security by design. Therefore, personal and non-personal data need to be treated with a high standard of data protection

Question 29. Do you know any technologies that can automatically differentiate personal and non-personal data within an end-to-end solution? (Open Question)

A summary of the different remarks that were provided:

- Very tricky to answer as there are still public debates about what is considered as a personal information or not. For example, if you can read the look of a person with his/her clothes. Is that considered as a personal information? There are also currently debates in the EU about the MAC address tracking with Wifi without having the consent of the user. Automation can only work if rules are very clear.
- Personal data are data that are attributed to one person or to a group of very similar persons. Here a hash function should be applied.
- Referring back to the data lake which should have a distinction layer between personal (sensitive) and non-personal data. This will also reflect in the data model which should capture this difference. So no technology but rather design of the setup.
- Non-personal data are in that respect aggregated data. Here it is not possible to trace back to a single person (e.g. “4711 persons went to the concert”). Here requirements for privacy are relaxed or even non-existing.
- An option is also to introduce a strict metadata layer that identifies all incoming data fields and has them flagged according to sensitivity. New data fields that are not present yet, are not handed to other systems but can be reviewed by an auditing process. Combined with a lineage graph of data transformations, graph analytics can be used to efficiently monitor these flows.
- Some automation tools, based on deep learning process, can flag and label structured and unstructured data so that it is possible to differentiate personal and non-personal data.
- To find personal data in Microsoft Azure, you can use tools like Azure Active Directory, Azure Data Catalog, Power Query (for Hadoop clusters in Azure HDInsight), Azure Search, and other related tools. Once you’ve found the personal data, you can use other tools such as Azure Information Protection to help implement data classification and apply persistent labels to the personal data.
- Microsoft PII (Personal Identifier Information). This Cognitive Skill is currently in public preview. This skill extracts personal information from an input text and gives you the option

of masking it. This skill uses the machine learning models provided by Text Analytics in Cognitive Services. More information can be found on these sites[16].

4.5.3 Other

Question 30. Do you have any suggestions / recommendations on the KPIs in the context of privacy & security? (Open Question)

A summary of the different remarks that were provided:

- the accuracy (miss-detection and false alarm) of the facie blurring system and license plate blurring system.
- The number of faces/license plates which were not blurred correctly per hour.
- Anonymize the private data direct at the data source and try to store as less private data as possible.
- Use of (jointly agreed) hash functions that cannot traced back to the single person. That function is to be applied as close to the sensor as possible (but it must be stressed that for some analysis that cannot apply).
- Adherence to a jointly agreed interface to the data base where NO back-traceable data is permitted.
- You should have a DPO assessing the solution as it will be developed once the architecture is put on paper.
- An automated platform measuring the elements of the Risk Assessment.
- Examples of KPIs relative to GDPR compliance:
    - Number of complaints including trends.
    - Number of PIAs/DPIAs.
    - -Number of DSR, categorized into each DSR (access, erasure, objection to processing, portability, etc.), including trends.
    - Number of data incidents.
    - Number of opt-ins where consent is necessary.
- Examples of KPIs relative to cybersecurity:
    - Intrusion attempts.
    - Mean Time Between Failures (MTBF).
    - Mean Time to Detect (MTTD).
    - Mean Time to Acknowledge (MTTA).
    - Mean Time to Resolve (MTTR).
    - Mean Time to Contain (MTTC).
    - Mean Time to Recovery (MTTR).
    - Cost per incident.
- Security: ISO standards for datacentres (see AWS / Google etc).
- The solution for privacy such as hash technologies will be critical for GDPR regulations.
- Continuous evaluation.

[16] How to use Named Entity Recognition in Text Analytics, Cognitive-search-skill-pii-detection , Protecting-personal-identifiable-information .

- The main indicator must be on the level of control of the end-user at every step of the flow from data fetching, to storing, processing and outputs sharing.
- System for monitoring the amount of sensitive data exchanged per period of time to warn of possible leaks.

Question 31. Do you have any specific privacy concerns that you would like to address? (Open Question)

A summary of the different remarks that were provided:

- The strategy/method of obtaining and managing consent for the pilot zone around Muntplein.
- It is very important to cluster the architecture and differentiate data collection from data mining/cleaning. The process of the data collection must be as short as possible and couldn't happen in the cloud => only metadata (highest form of anonymization) is transferred to the cloud.
- Uncertainty about the privacy shield.
- Think of the ability to delete personal data, including backups, in the context of a GDPR request. E.g. data related to the mobile phone.
- The GDPR is not a constraint but an opportunity to put the end-users in control of all their personal data and be able to extract the value of it in a user-centric way, not an organisation centric one. A centralised architecture and data flow will therefore probably not work. Hence, it would be interested to challenge the market to implement a more ambitious decentralised approach, involving at every stage the end-users and approach.

4.6 Data quality

Regarding the assessment of the data quality, the Muntstroom PCP groups will request data quality management plans from the technology vendors and their compliance thereof. But maybe there are other options.

Question 32. How would you ensure the quality of your data? How can this be evaluated? What are the KPIs? (Open Question)

A summary of the remarks that were provided:

- A well-designed data quality management plan which includes measurable KPIs and regular reports and deliverables will be sufficient to ensure data quality. Moreover, external data audits can be done at various moment in time.
- Examples of topics that could be included are:
    - A data checking algorithm in the processing/storage layers. These algorithms will check different aspects of the data quality and modify it consequently.
    - A description of the type of data which is being collected, for which use case.
    - The usage of data sources that can measure the same metric.
    - Data sources certification: integrity of raw data, no alteration in the fetching process and ETL.
    - The need for a visualisation component to monitor data quality (data loss, gaps in data streams, accuracy indicators, inconsistencies alarms, etc).
    - By Automated validation procedures, explicit accuracies (qualitative by metadata, quantitative by confidence intervals).
    - Manual tests and checks (accuracy measurements).
    - Two types of quality control:
        - Technological: the format, update frequency, connection time, errors,..
        - Content: false positives, false negatives, compare with manual counting,..
    - The need for a preprocessing approach to evaluate the data quality to ensure that data has no noise or missing values.

<u>Possible KPIs</u>

- Integrity: there is no loss of data and it remains relevant.
- Consistency: the data is uniformly coherent.
- Completeness: there is no missing data.
- Ratio of Data to Errors.
- Number of Empty values.
- Data Transformation Error rates.
- Data Storage Costs.
- Data-time-to-value.
- Duplicates of Data.

4.7 Financial aspects

During the PCP technology providers need to design the solution (PCP-phase 1), build a prototype (PCP-phase2) which will be tested during the Living Lab (PCP-phase 3). The Muntstroom PCP group would like to know which financial involvement is needed to co-fund these activities.

4.7.1 The needed financial involvement

Question 33. How much do you estimate it will cost you in total to design the solution (PCP-phase 1, in €)

More than half of the respondents did not answer. E.g. because they found it difficult to make an estimation without a predefined scope, without knowing the amount of use cases and/or without having found the right partners.

The estimates range from several € 10k until € 160k.

The average estimation of 15 respondents is about 58k.

These estimates were based on e.g.

- The geographical scale of the pilot.
- The duration as published .
- Running a full commercial programme team (including risk, change, communication, security and GDPR expertise and management).
- License costs.
- Hardware and potential associated taxes, VAT...

Question 34. How much do you estimate it will cost you to develop the prototype (PCP-phase 2, in €)

More than half of the respondents did not answer. E.g. because they found it difficult to make an estimation without a predefined scope, without knowing the amount of use cases and/or without having found the right partners.

The estimates range from € 25k until € 350k.

The average estimation of 13 respondents is about 165k.

These estimates were based on e.g.

- The geographical scale of the pilot.
- The duration as published .
- Running a full commercial programme team (including risk, change, communication, security and GDPR expertise and management).
- License costs.
- Hardware and potential associated taxes, VAT...

Question 35. How much do you estimate it will cost you to install and operate the Living Lab (PCP-phase 3, in €)

More than half of the respondents did not answer. E.g. because they found it difficult to make an estimation without a predefined scope, without knowing the amount of use cases and/or without having found the right partners.

The estimates range from € 40k until € 375k.

The average estimation of 17 respondents is about 170k.

These estimates were based on e.g.

- The geographical scale of the pilot.
- The duration as published .
- Running a full commercial programme team (including risk, change, communication, security and GDPR expertise and management).
- License costs.
- Hardware and potential associated taxes, VAT...
- The small number of sensors. Where the cost per sensor/use case diminishes exponentially with implementation scale.

Question 36. What would your estimated total Research & Development cost be? (in €, PCP phases 1, 2 and 3)

More than half of the respondents did not respond. E.g. because they found it difficult to make an estimation without a predefined scope, without knowing the amount of use cases and/or without having found the right partners.

The estimates range from € 40k until € 800k.

The average estimation of 13 respondents is about 330k.

Question 37. Could you give a rough overview of the highest, possible costs that you would make?

	Costs in €
a. use of sensors / hardware / … (e.g. rental costs)	2k, 60k, 25k for 6 months renting of 10 thermal cameras + communications devices + other hardware
b. use of software	1k, 18k, 12k,
c. installation costs	2k, 24k, 30k
d. operational cost (including energy consumption)	200/month, 8k/year, 2,5k (basically communications)
e. maintenance costs	100/month (server), 9k/year, 7,5k
f. other, like ….	

The most important costs are probably:

- Initial cost of hardware (camera, chipsets) and installation fees are the most important involved costs.
- Time for development resources, and time for maintenance of the software stack.
- Installation and maintenance costs seem to us at this stage to be the most important costs.
- R&D and use of platform/software.
- Data collection.
- Installation and data preparation phase + model training and implementation costs will be the highest.
- Renting of the IoT datahub.
- Using early existing cameras could reduce the CAPEX.

Question 38. Could you reason on the financial pros and cons of using Open Source or non-Open Source components/software? (Open Question)

Pros	• Open Source fosters collaboration in the ecosystem (data providers, researchers, cities, governments, smart city platforms), which is what you need when developing the desired end-to-end solution. • You can have access to the code and then it is easier to integrate that tool to others and you can analyse the code to know what it is behind the tool (ethical issue). • Using Open Source Software yields a lower total cost of ownership when compared to closed source and proprietary alternatives. Adopting Open Source Software generally has a lower up-front cost (because the software often comes at no cost or relatively low cost) and shifts the cost centre from licensing (an operating expense) to customization and implementation (a capital expense). Additional costs like training, maintenance, and support are sunk costs. You're going to be paying for both regardless of if the software is open or closed source, the cost often being baked into the license in the case of commercial, off-the-shelf software (COTS). What makes Open Source unique is that you're not paying for the right to use the underlying intellectual property[17]. • The fact that they are usually platform agnostic, makes them suitable for use in numerous applications. As they are not intended to be exploited, they do not integrate policies and methodologies that can lead on a vendor lock-in situation. • An open standard guarantees a high level of interoperability. It is easier to use for the greatest number of people and this makes innovation emerge more quickly.
Cons	• The overall package for developing an end-to-end must be commercially interesting for the participants. As they are taking a business risk. • Open Source material could have limitations in the commercial exploitation. Also, you would only rely on the community for specific developments. • OpenSource doesn't always mean cheaper, especially since Open Source can incur more development costs. Hence, the use of Open Source with enterprise support is often desirable. This allows to start without any cost but when OPEX cost becomes too large, part of that cost could be shifted to CAPEX by moving it into a different service contract with a provider. • The fact that an Open Source Software has its source code open can reveal vulnerabilities, that may be exploited by developers with bad intentions. Security threats, especially in a service, that will make use of personal data, even if anonymized, shall be greatly taken into consideration. • The fact that are not aimed for commercial use, has as a result in most of the cases a less user-friendly environment. • For Open Source technologies the learning curve maybe higher. • OSS is more sensitive to data breaches.

[17] https://opensource.com/life/15/12/why-open-source

Question 39. Could you give a rough overview of the highest, possible costs that you would make in relation to the proposed architecture:

	Costs in €
a. Physical layer	20k, 72k
b. Communication layer	12k, 5k[18]
c. Storage layer	8k
d. Processing layer	80k
e. Advanced analytics layer	150k, 25%, 180%
f. Smart Access layer	45k
g. End-user layer	42k

Table 4: Expected costs per architecture layer

	Anwers
a. Physical layer	1
b. Communication layer	
c. Storage layer	
d. Processing layer	6
e. Advanced analytics layer	6
f. Smart Access layer	3
g. End-user layer	1

Table 5: Expected most expensive layers (multiple answers given)

[18] Includes API to connect with the existing upper layers.

4.7.2 Buffer times in the Living Lab-phase

In Muntstroom PCP-phase 3 (Living Lab) a prototype will be tested in a Living Lab. This will require the installation of the system, a start-up period to finetune the settings and a learning period before analytics can function in full.

Question 40. How much time do you think the suggested buffer times could be (in weeks)?

	Amount of weeks
a. Start-up period to a. finetune the setting / do the calibration	From 2-3 weeks to 8 - 14 weeks
b. Learning time for descriptive analytics	From 2-4 weeks to 10 - 20 weeks
c. Learning time for the deeper analytics	From 4--8 weeks to 10 - 20 weeks

Additional remarks:

- Are these the learning times per use case?
- Time to reach higher confidence levels is highly depending on the exact scope, concept, architecture, data availability and HW used. With the use of 3D-sensors, this can be super-fast. However, a relatively high amount of time should go into right positioning and calibration of sensors.
- The buffer times indicated are for a certain period in the year. It would be wise to extend that period to a longer period if one wants to have enough people flow examples on different moments and seasons of the year. For instance, summer time = more tourists and cultural events, while in other seasons (strong winter and very cold) people flow will be different. Training data should be as representative as possible to cover all cases/seasons/conditions.

Question 41. What would be the personnel costs of each of the suggested buffer times (in man-days)?

	Personnel costs in man days
a. Start-up period to a. finetune the setting / do the calibration	20, 30, 45, >50, 20, 60, 150 - 250
b. Learning time for descriptive analytics	20, 15, 30, >25, 20, 80, 100- 200
c. Learning time for the deeper analytics	24, 30, 60, >25, 20, 160;100 - 250

Additional remarks:

- 30-40 man days/use case, whether that's more descriptive or advanced analytics.
- This is too complex to answer in this state of the project. Especially since it is not yet clear which data sources can be used.

4.7.3 Co-funding

The Muntstroom project gives technology vendors the possibility to pilot solutions that can be later commercialised to public / private buyers. This also allows helps to speed up the time-to-market of these solutions. Therefore, there are different incentives for technology vendors to co-fund the R&D together with the Muntstroom PCP Group.

Question 42. How much of your total R&D-costs (see the question before) would you be willing to fund yourself? (With a minimum of 20%)

	Answers
20	7
25	5
30	3
40	1
50	2
60	1
No anwer	20

4.7.4 Royalty scheme

In case the developed proprietary solution can be successfully commercialized to third parties, the Muntstroom PCP Group wants to set up a fair pay back scheme of their investments done in the PCP.

Question 43. Are you familiar with the concept of a royalty scheme approach?

	Answers
Yes	21
No	10
No Answer	8

Question 44. Are there any suggestions and/or recommendations you can make in order to set up such as fair pay back scheme?

E.g. regarding the maximum percentage of the PCP investments you would consider as a fair amount to pay back to the participating contracting authorities, the maximum duration of the pay back scheme, the support you would expect from the contracting authorities (by acting as a reference project) in commercializing the developed proprietary solution to third parties, etc.

A summary of the different remarks that were provided:

- An alternative way would be to split the project into 2 parts:
  - A common Open Source solution => free of charge.
  - A "commercial & consultancy" solution => paid for.
- Under the scope of the travel and security related work packages mainly existing solutions (even when they are currently in the development phase) are applied. That means that the R&D is spent by the companies and Muntstroom will not receive royalties.

- Moreover, the Phase 3 winning consortia will try to bid with best pricing, the partial support awarded in phase 3 will be compensated by a lower cost allocation in this quote towards phase 4.
- Based on API usage and requested computing power (Software as a Service).
- If a fair pay back scheme is set up, the Muntstroom PCP Group should do more than just be a reference project. Your network of potentially interested parties will probably grow throughout the PCP project as well, creating interesting leads for a commercially viable product, once this would be finished. Having a good reference case is one thing, having that reference case speak out and play an active role in marketing & selling the solution is something else.
- E.g. regarding the maximum percentage of the PCP investments you would consider as a fair amount to pay back to the participating contracting authorities, the maximum duration of the pay back scheme, the support you would expect from the contracting authorities (by acting as a reference project) in commercializing the developed proprietary solution to third parties, etc.
- If the solution of Muntstroom PCP project is successfully commercialized, one could propose a %-profit returns to the participating contracting authorities until the initial investment is paid off, plus two years after, due to the benefits that will arise from this a co-operation such as the referral marketing and brand awareness.
- You could develop a business plan to evaluate different royalty schemes (pay per use, subscription,…) that depend on different factors: data volumetry, type of clients, open data/data brokers,… but also maintenance costs.
- A royalty should be levied on the turnover generated by the data resale platform.
- The pay-back period is to be agreed and then multiple simulations could be done to conclude on the best options (taking into account different hypotheses and assumptions).
- Advertising the case and platform is the most important role for the contracting authorities, on the way to further commercialise the proprietary solution. Much lower costs on support and maintenance, free delivery of extensions required by other "customers" are the benefits to anticipate within such a schema.
- Use a billing engine for both commercial data sharing and revenue sharing.
- The Epic architecture is highly modular. So it is easy to earmark which components are subject to licenses.
- Multiple scenarios are possible. Royalties as a percentage of revenue, decreasing over time. After a while, the contractor does the business development effort, the operational deployments, the support, the maintenance and the roadmap evolutions,...

4.8 PCP challenges and complexity

All the above challenges and questions will be brought together in one PCP-process.

The Muntstroom PCP Group uses a PCP-approach to procure the required innovation. See the figure below. A PCP is a competitive and risk sharing procedure where technology vendors develop solutions in subsequent phases. See for more information the market consultation document.

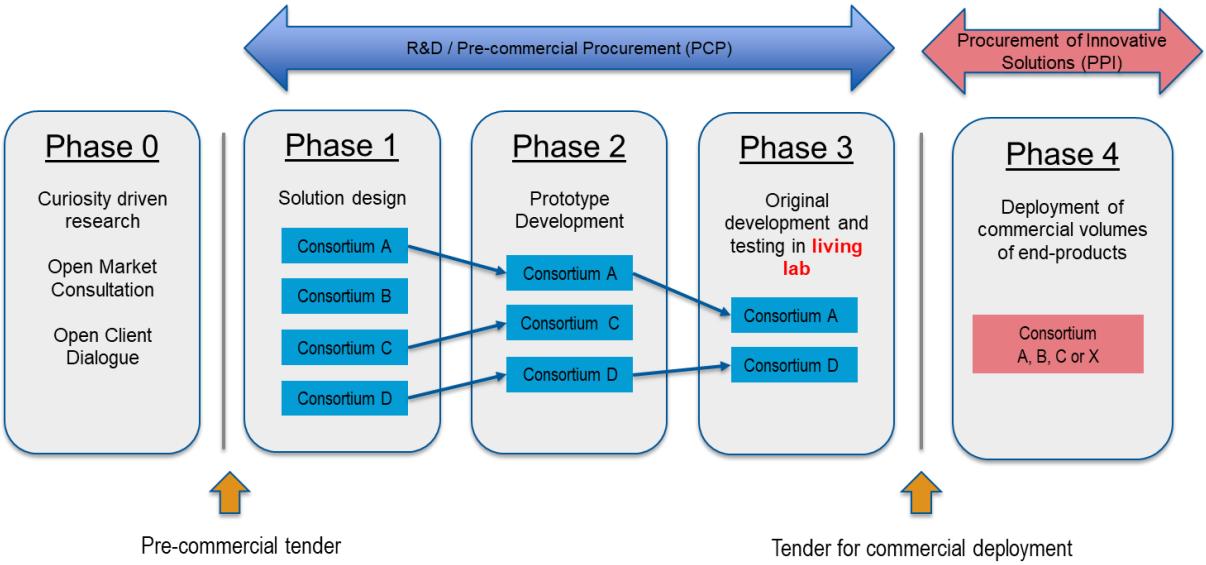

Figure 2: the different phases of Innovation procurement

Question 45. Are you familiar with the concept of Innovation Procurement / Pre-commercial procurement?

	Answers
Yes	18
No	11
No Answer	10

The Muntstroom PCP group proposes the following duration of the 3 PCP-phases:

1. *PCP-phase 1 (solution design): 3 months.*
2. *PCP-phase 2 (prototype design): 7 months.*
3. *PCP-phase 3 (test in a Living Lab): 6 months.*

Question 46. Is the proposed timeline a good estimation, based on the complexity and scope of the project? If not, please elaborate. (Open Question)

A summary of the different remarks that were provided:

Regarding the timing

- For about half of the respondents the first estimation seems reasonable. Although the planning does depend heavily on the amount of use cases that have to be shown in phase 3 and seems to follow the scenario "that everything goes according to plan".
- Different respondents find the planning too ambitious and suggest some additional months per phase.
- One supplier claims to have a solution in TRL 6 and suggests to speed up the process to phase 1 of 1 month, phase 2 of 2 months and phase 3 of 6 months.

General remarks

- Yes, if the scope is severely limited: use case #1 (basic building block) + a few additional use cases.
- The challenge is to ensure that the consortiums and people are in place before the stated time.
- You could already open the Living Lab already during the prototype design phase. It will be important during the prototype design phase to be able to check on some assumptions for real-life testing in the Living Lab.
- The objectives of the project are too many and too ambitious. Our suggestion is to scope the project in different parts (and maybe different PCPs) with different KPIs to measure ROI and prioritize the different use cases following technical complexity/business requirements. In this way the use of Design thinking approach and model canvas can succeed. An adequate timeline could be evaluated during this process.
- For instance thermal cameras are costly devices. You could suggest some longer test period in order to increase the results/deployment costs ratio otherwise the renting price of the system and of the service can result too high.
- The Phase 2 Prototype should be more targeted than experimental.
- The solution Design is the most important part. But highly depending on the scope split and experience of consortium partners. You could start small, with more time, and evolve the concept and the system over time.
- The planning is slightly underestimated because A.I.is very much a trial and error approach. There will be a lot of coordination, back and forward discussions, evaluations, adaptation of the approach, testing, feedback, review of models etc.

Question 47. Do you see the objective of the Muntstroom PCP-project (developing and testing an end-to-end solution for People Flow) as too complex or too long-winded? If so, which part? (Open Question)

A summary of the different remarks that were provided:

- This is the first time we are involved in such a PCP and it looks like an innovative approach/project. We can't say yet on how to compare.
- The complexity and potential duration of the project are commensurate with its ambitions: dense, challenging, exciting.
- Not really – to the contrary, it should be seen as a pilot to roll into production once proven. Then you could also consider general traffic.
- We are concerned that the focus of the WPs is often “too open”, i.e. there will be many discussions about the scope definition and not enough about the solution.
- We recommend keeping it simple for the first release and to develop the systems in several steps.
- The objectives of the project are clear and easy to grasp. However, a narrower approach of the use cases, where additional features and AI-technologies are added in further stages of the development, could yield better results.
- A longer test period, e.g. one year, would permit having four seasons results, which would make results more representatives.
- The current scope of the project with the large number of use cases is far too complicated. It would be better - in my opinion - to try to build an end-to-end solution (i.e. with the complete architecture and all application layers in place) for a restricted number of basic analytics modules (e.g. use case #1, 'understanding the pedestrian'). Additional use cases can be added later on, perhaps using the same PCP approach.
- The challenge is complex and will a priori only be possible through multi-disciplinary consortiums.

Question 48. What do you see as the main challenges from your side? (one answer possible)

	Answers
a. The integration of individual components (e.g. sensors, data-storage and analytics) to an integrated end-to-end solution.	6
b. The integration of the business cases of individual components into one integrated business case to deliver the end-to-end solution.	10
c. The development of the Analytics-layer	5
d. The development of the Smart Access-layer	3
e. Other, namely ….	7
No Answer	8

Other, namely:

- The privacy and confidentiality of data (in conjunction with consent & permissions in living lab phase).
- GDPR compliance and Privacy Shield.
- Data standardization and support for a large number of sensor types (and anticipation of types which can be added in the future).
- Prioritisation of use cases and determination of feasibility.
- Technically we have proven references. We are convinced the challenge will be to setup a framework in relation to privacy.
- Finding and connecting the right / most suitable partners to provide the best smart solutions for all use cases.
- Defining the initial success factor for Muntstroom that will still hold after implementation.

4.9 Last remarks

Question 49. Any last remarks? or would you like to upload a file?

- The PCP-method is new for us. So we have answered the questions as much as possible. We expect to have much more details to share as the PCP projects proceed.
- We are extremely interested in the project, in its perspectives, ambitions and context; we have already been able to get in touch with other Belgian companies/start-ups, whose added value we are convinced of, on the one hand for the Muntsroom project, but also on the other hand, in the framework of a collaborative partnership.
- The synergies resulting from the work between our 3 entities seem to us to be extremely complementary and promising, given their respective expertise. We are therefore planning to respond in a unified way, in the form of a consortium, to the upcoming call for tenders, and have already exchanged in this sense several times before the 4/12 deadline (for greater cohesion and ease at this stage it has been agreed that each entity should fill in a questionnaire).
- After taking into account the elements relating to the legal, ethical and sometimes even philosophical aspects of the project, we called upon our close contacts, active on these subjects within a university. We are convinced that their approaches could be more than useful, essential, in the design of any envisaged solution, and could therefore call upon them in the framework of this possible consortium to be set up.
- Depending on the exact scope of the Phase 3 prototypes and tests, the foreseen budgets might be insufficient. Installing onsite equipment has proven to be very complex and expensive. There is a need to first develop a joint vision and become clear about the exact scope. Next steps should focus on "paper result" not on building prototypes with an unclear scope definition.

5 Summary results OCD-questionnaire

This chapter summarises the results of the Open Client Dialogue. This market consultation was targeting companies that would like to use the -to be produced- People Flow-data to develop new apps/services for commercial purposes.

5.1 About the respondents

In total 8 interested parties replied to the questionnaire. The general findings and results are summarised in the following sections.

5.2 The interest in People Flow

Question 1. Why would you be interested in People Flow big data (analytics)? (Maximum 5 answers possible) .

	Answers	Ratio
Stimulate the use of combined mobility	7	87,5%
Improving accessibility	5	62,5%
Support event management	4	50%
User-oriented urban planning / development	3	37,5%
Support safety and security	3	37,5%
Promotion of the mode walking	2	25%
Commercial potential / Marketing	2	25%
Asset management	2	25%
Guidance for (foreign) visitors	1	12,5%
Others, namely	0	0%
No Answer	0	0%

Table 6: Answers ordered by amount of answers

Question 2. Do you already collect People Flow-data yourself? (Yes/no)

	Answers
Yes	4
No	4
No answer	0

Yes, for instance as a data marketplace or a manager of security cameras.

5.3 Participation in the PCP

During the Muntstroom PCP a technology will be developed and tested to visualise people flow (counting, direction, speed), facilitate joint big data-analytics and make the People Flow-data available for a wide array of users.

Question 3. Are you interested in participating in a Muntstroom community (eg. presentations and workshops about the use of People Flow-data) throughout this project?

	Answers
Yes	8
No	0
No answer	0

Question 4. Are you interested in testing the Open Data during the living lab phase?

	Answers
Yes	8
No	0
No answer	0

5.4 Use Cases

To test in the Living Lab-phase if the developed solution meets the expectations, use cases have been drawn up. See Annex 2) Use cases.

Question 5. Could you prioritise the analytics-use cases in order of importance to you? (Maximum 5 answers possible. Score low 1 – 3 high importance)

Use case	1	2	3	No answer
1. Pedestrian analytics	2	2	2	2
2. Analysis of the reasons for movements	1	3	2	2
3. Event analytics	1	1	4	2
4. Crowd analytics: security	1	3	3	1
5. Crowd analytics: COVID	0	3	4	1
6. Mobility hub analytics: Usage of specific public transport lines	2	0	3	3
7. Mobility hub analytics: Transfers via gates, platforms and corridors	2	1	3	2
8. Asset management	2	2	2	2
9. Commercial analytics: shopping policy and research	2	0	4	2
10. Commercial analytics: individual shop	2	1	3	2
11. Traffic lights analytics	2	2	2	2

Table 7: Answers ordered by number use case

Use case	1	2	3	No answer
5. Crowd analytics: COVID	0	3	4	1
3. Event analytics	1	1	4	2
9. Commercial analytics: shopping policy and research	2	0	4	2
4. Crowd analytics: security	1	3	3	1
7. Mobility hub analytics: Transfers via gates, platforms and corridors	2	1	3	2
10. Commercial analytics: individual shop	2	1	3	2
6. Mobility hub analytics: Usage of specific public transport lines	2	0	3	3
2. Analysis of the reasons for movements	1	3	2	2
1. Pedestrian analytics	2	2	2	2
8. Asset management	2	2	2	2
11. Traffic lights analytics	2	2	2	2

Table 8: Answers ordered by "high importance"

Question 6. Could you prioritise the routing-use cases in order of importance to you? (Maximum 4 answers possible).

	1	2	3	No answer
a. Routing PMR: wheelchair	1	2	4	1
b. Routing PMR: partially sighted	1	4	2	1
c. Indoor routing: subsurface	3	2	2	1
d. Indoor routing: building	3	1	3	1

Table 9: Order by number use case

	1	2	3	No answer
a. Routing PMR: wheelchair	1	2	4	1
d. Indoor routing: building	3	1	3	1
b. Routing PMR: partially sighted	1	4	2	1
c. Indoor routing: subsurface	3	2	2	1

Table 10: Answers ordered by "high importance"

Question 7. Do you have any remarks on any or several of the use cases?

Use case	Remarks
Pedestrian analytics	It would also be interesting to understand pedestrian accidents.
Analysis of the reasons for movements	The bottleneck for this use case is probably not the technology but the backhaul connection and HPC.
Crowd analytics: COVID	This is already running on the Meir, Antwerp.
Mobility hub analytics: Usage of specific public transport lines	Bonus and Basic seem reversed. Large boarding movements are possible to detect but will require larger data sets.
Asset management	Maybe add the detection of Points/areas of Interest (POIs) and of common trajectories/routes. And how these may change in cases of specific events and/or because of weather conditions.
Commercial analytics: shopping policy and research	This use case, based on the data input expected, could be further expanded to have predictive and prescriptive analytics (adding to the difficulty) in order to increase business value. Assuming the use of additional datasets/information, e.g. on events and weather, further correlations can be detected.
Traffic lights analytics	A smart crossing is already running in Antwerp

5.5 Your needs

The Muntstroom PCP Group has defined three general outputs of the Muntstroom solution:

a. *Data via Open Data (limited data sets) and Broader data sets (shared and closed data)*
b. *Visualising people flow (the amount of pedestrians, the direction in which they walk and their speed) in Maps and graphics.*
c. *Supporting people flow (facilitate positioning & routing).*

The questions in this block are about aligning these desired output with your needs.

5.5.1 People-flow metrics of your interest

The public buyers intend to require structuring of the collected data in accordance with the Public Life Data Protocol (2017) of Gehl Institute.

Question 8. Are there specific metrics that you are interested in?

	Answers
a. The subcategories (Walking, Running, Supported, Carried, Rolling)	3
b. Advanced content (more details within the subcategories)	2
c. No specific metrics	3
d. Other, namely…	0

Question 9. Can you indicate whether there are official/other People Flow-standards that could be (more) suitable for the Muntstroom PCP? (Open question)

No answers to this question were received.

5.5.2 Your data needs

Question 10. How can the data help you get where you want to go? E.g. a specific data set could help you to answer a certain question. Or a combination of data sets that you hope will deliver a certain insight.

- People flow data, especially if delivered in high granularity, can help deliver important people flow analytics. However, it is the combination of multiple datasets that can bring business value in the use cases and can provide new insights a human might not see. This better understanding of how people move can help us provide better quality information about the state of alternatives mobility networks, how pedestrians are using public space and help preventing incidents and accidents.

Question 11. What question(s) align best to your interests? (Multiple answers possible)

	Answers
a. What happened? (in retrospect/looking back)	7
b. Why did it happen? (introducing causality)	3
c. What will happen? (prediction of the future)	7
d. How can we make it happen? (scenario planning)	3
e. No answer	0

Question 12. What kind of output do you need (multiple answers possible.)

	Answers
a. Raw data (I would like to analyse this data myself)	8
b. Processed data (maps and graphics on counting, direction, speed)	5
c. Data insights (Insights generated by drawing conclusions from processed, analysed data)	4
d. No answer	0

5.5.3 Your data requirements

Question 13. What kind of data (sets/formats) do you require? (Open Question)

- Different formats were mentioned (CSV, JSON, APIs,…)

Question 14. What frequency do you expect?

	Answers
a. Open access 24/7	6
b. Data dump / week	1
c. Data dump / month	1
d. Data dump / year	0
e. Other, namely…	0
f. No answer	0

- "Ideally, open access 24/7, but for specific purposes weekly data dumps could also be useful."

Question 15. Which granularity would you like? (different possibilities possible)

	Answers
a. Per minute	6
b. Per hour	3
c. Per day	2
d. Per week	0
e. Per month	0
f. All	0

Question 16. Do you require…?

	Answers
a. Historical data	2
b. Real Time data	0
c. Both	6
d. No answer	0

Question 17. Do you have specific requirements regarding the related data quality? (Open Question)

- It would be good to have information related with the positioning accuracy of the data offered, as well as (for analytics-based results) the performance metrics of the methods used.

Question 18. Do you have additional requirements that have not been mentioned before? (Open Question)

- A detailed documentation for the type of data offered and related APIs would be useful.
- Being able to have origin-destination insights would be of great help!

5.5.4 Learning time

Before using new data /a new technology to the full extent, normally a certain learning time is taken into account.

Question 19. In order to be able to use the new application faster, do you have any wishes? (Multiple answers possible)

	Answers
a. The possibility to discuss the content of the captured data	4
b. An explanation workshop on the technology	6
c. A regional knowledge community around People Flow-data analytics	4
d. Other, namely	0
e. No answer	0

5.5.5 Output requirements

Question 20. What kind of type of API would you require?

- Given documentation, we can adjust to different choices. Ideally an open-data API in any standard format.

Question 21. Which demands do you have in relation to the Open Data-output? (Multiple answers possible)

	Answers
a. I am not interested in Open Data.	0
b. API (I will do the data crunching myself).	8
c. Visualisation of the People Flow-data. Like in Melbourne[19]	3
d. Visualisation of the People Flow-data in relation with other Open Data sets. Like in Chicago[20].	2
e. Other, namely….	0
f. No Answer	0

[19] http://www.pedestrian.melbourne.vic.gov.au/

[20] https://chicago.opengrid.io/opengrid/

5.6 Value of Data

The foreseen Muntstroom solution is supposed to have two kinds of data output:

1. *Limited data sets (Open Data)*
2. *Datasets on demand (Shared and Closed data)*

Question 22. When would "Datasets on demand" have a higher value for you than basic "Open Data"? (Multiple answers possible)

	Answers
a. Broader data sets (higher frequency, granularity, …).	4
b. Specific data sets, if so which? See the following question).	5
c. Combined with other data sets, if so which (see the following question).	4
d. Real time data.	5
e. I have no idea yet.	0
f. No Answer	0

Question 23. Which specific data would you be interested in? (Multiple answers possible)

	Answers
a. A specific moment of the day / week/month/year.	5
b. During a specific event.	8
c. Searching for a particular incident / occurrence.	5
d. By type of asset.	3
e. Per speed of pedestrians (e.g. strolling shoppers).	2
f. Data per sensor (data of a certain area).	5
g. Real time Data.	4
h. Other, ...	0
i. No idea yet	0
k. No answer	0

Question 24. With which other data sets would you like to combine the People Flow-data? (Open question)

- Map information (e.g. on shops, landmarks), information on events and weather conditions, additional metadata on shops (i.e. descriptions, opening/closing hour, etc).
- Examples could be GIS data, google maps, anonymized mobile data set (commercially available already).
- Air quality, weather, available micro-mobility around.

5.7 Privacy and security

The solution should follow the privacy and security-by-design and by-default principles and the output data will therefore be completely anonymised.

Question 25. Do you have any specific privacy concerns you would like to address? Are there e.g. combinations of data that could create privacy concerns? (Open question)

- Depending on the use case, think next to GDPR also of the camera law.
- What consent model(s) are being proposed/used?

Question 26. How can we create more trust between organisations to stimulate data sharing? (Multiple answers possible):

	Answers
a. Promotion of a government-to-government data-sharing culture.	3
b. Promotion of a business-to-government and government-to-business data-sharing culture (for the public interest).	4
c. Transparency of any data-sharing collaboration.	7
d. An established safe and ethical data-sharing mechanism.	7
e. A client board to discuss complaints.	3
f. Other challenges, namely…	2
g. No Answer	0

5.8 Last remarks

Question 27. Any last remarks? Or do you want to upload a document?

Good luck!

6 Questions and answers

During the market consultation process questions were raised by email and via the chat during the two webinars. A Q&A summary was made in French, Dutch and English.

The Q&A summary can be downloaded via this link (click here).

7 Conclusion and follow up actions

7.1 Conclusion

From desktop study it emerged that:

1. Elements of the desired end-to-end solution do already exists, but the solution as such not yet.
2. There is room for innovation in the fields of
    > Analytics software.
    > Big data analytics software.
    > The integration of different existing elements into one integrated total product.

The results of the Open Market Consultation show that the market parties support this conclusion. Only one party mentioned that they have an end-to-end solution of TRL[21] 6 (technology demonstration), where others (at least 31 out of 39) are still looking for partners to integrate their developed elements with that of others, in order to develop and test the desired end-to-end solution.

The results of the market consultation therefore lead to the conclusion that there is room for starting a PCP, based on TRL 6.

7.2 Follow up actions PCP and PPI

Intention to launch a PCP and PPI

As a follow up of the market consultations, it is the Muntstroom PCP Group's strong intention to define and set up a Pre-Commercial Procurement of R&D services. The elaboration of the PCP tender documents will take into account the analysis of the feedback of the market parties.

It is also intended that after successful completion of the Muntstroom PCP, a follow-up Public Procurement of Innovative Solutions (so-called PPI) will be initiated, based upon the lessons learned from the Muntstroom PCP.

Year	Date	Activity
2021	Q1	Start matchmaking activities via private sector intermediates
		Webinars on the theory of a Pre-Commercial Procurement
	Q2	Foreseen start of the PCP via a pre-commercial tender
2022	Q1-2	Preliminary start of the Living Lab-phase (PCP-phase 3)
	Q 3-4	Preliminary start of the PPI (phase 4) via a tender for commercial deployment

Table 11: Preliminary timetable of the intended PCP

[21] Technology Readiness Level.

List of Acronyms

Abbreviation	Word
BCR	Brussels Capital Region
CIRB – CIBG	The Brussels Regional Informatics Centre
CMS	Content Management System
Eafip initiative	European Assistance for Innovation Procurement initiative
EU	European Union
FAIR	FAIR-principles: Findable, Accessible, Interoperable, Re-usable[22]
FRAND	Fair, Reasonable and Non-Discriminatory
IPR	Intellectual Property Rights
MaaS	Mobility as a Service
OECD	The Organisation for Economic Co-operation and Development
OCD	Open Client Dialogue
OMC	Open Market Consultation
OSS	Open Source Software
PA	Process Automation
PCP	Pre-Commercial Procurement
PETs	Privacy Enhancing Technologies
PPI	Public Procurement of Innovative solutions
PRM	Persons with reduced mobility
R&D	Research and Development
SEP	Standard Essential Patents
SOTA	State of the Art Analysis
STIB-MIVB	Brussels Intercommunal Transport Company
TCO	Total Cost of Ownership
TED	Tenders Electronic Daily
TRL	Technology Readiness Levels

[22] Directive (Eu) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. Official Journal of the European Union L 172/56, 26.6.2019. Recital 27. Article 10. Research data. *"1.Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'), following the principle of 'open by default' and compatible with the FAIR principles. In that context, concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests, shall be taken into account in accordance with the principle of 'as open as possible, as closed as necessary'. Those open access policies shall be addressed to research performing organisations and research funding organisations."*

List of Definitions

Definition	Explication
Belgian Public Tender Bulletin	Bulletin des Adjudications/ Bulletin der aanbestedingen. Website: https://enot.publicprocurement.be/enot-war/
Big data	Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations.
Client	A person who receives services. Potential public and private parties who may use and re-use people flow related open and shared data in the Brussels Capital Region.
Data-driven insights	Insights generated by drawing conclusions from processed, analyzed data.
Internet of Things	Internet of Things (IoT) is a network of Internet connected objects/devices able to collect and exchange data.
Muntstroom PCP Group	The joint public buyers behind this PCP, being public transport operator STIB-MIVB, Brussels Regional Informatics Centre CIRB-CIBG, the regional authority Brussels Mobility and regional agency Parking Brussels.
Open data[23]	Open data are digital data that have the technical and legal characteristics required to make them freely available for use, re-use and republish at anytime and anywhere, without restrictions from copyright, patents or other mechanisms of control. Data quality is one of the factors that affect the successful use of open data.
Open data licenses	Standard licenses that are available online, which allow data and content to be freely accessed, used, modified and shared by anyone for any purpose, and which rely on open data formats. The re-use of documents should not be subject to conditions. However, in some cases justified by a public interest objective, a license may be issued imposing conditions on the re-use by the licensee dealing with issues such as liability, the protection of personal data, the proper use of documents, guaranteeing non-alteration and the acknowledgement of source. If public sector bodies license documents for re-use, the license conditions should be objective, proportionate and non-discriminatory. [24]

[23] Directive (Eu) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. Official Journal of the European Union L 172/56, 26.6.2019. Recital 16)

[24] Directive (Eu) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. Official Journal of the European Union L 172/56, 26.6.2019. Recital 44. *(…) Member States should encourage the use of open licences that should eventually become common practice across the Union. Without prejudice to liability requirements laid down in Union or national law where a public sector body or a public undertaking makes documents available for re-use without any other conditions or restrictions, that public sector body or public undertaking may be allowed to waive all liability with regards to the documents made available for re-use."*

PCP	Pre-Commercial Procurement (PCP) is a specific approach to procure R&D services that involves competitive development in phases, risk-benefit sharing under market conditions, and where there is a clear separation between the PCP and the deployment of commercial volumes of end-products (potential follow-up PPI).[25]
Public buyers	The joint public buyers behind this PCP, being public transport operator STIB-MIVB, Brussels Regional Informatics Centre CIRB-CIBG, the regional authority Brussels Mobility and regional agency Parking Brussels.
Privacy Enhancing Technologies	Technologies like anonymisation and encryption techniques, protocols for anonymous communications, at-tribute based credentials and private search of databases.
Pre-processed data	Pre-processing includes, for example, cleaning, instance selection, re-sampling, normalisation, transformation, feature extraction and selection.
Shared data	Data that is shared with a certain group of users that have specific access rights for particular purposes.

[25] See eafip Toolkit, at www.eafip.eu

Annex

Annex A. Visualisation of desired solution

To shape the pedestrian monitoring system, the general idea is to design and test a system that is designed to 1) capture, 2) communicate, 3) store, 4) process 5) analyse and 6) provide smart access to people flow data. These aspects are visualised in the figures below.

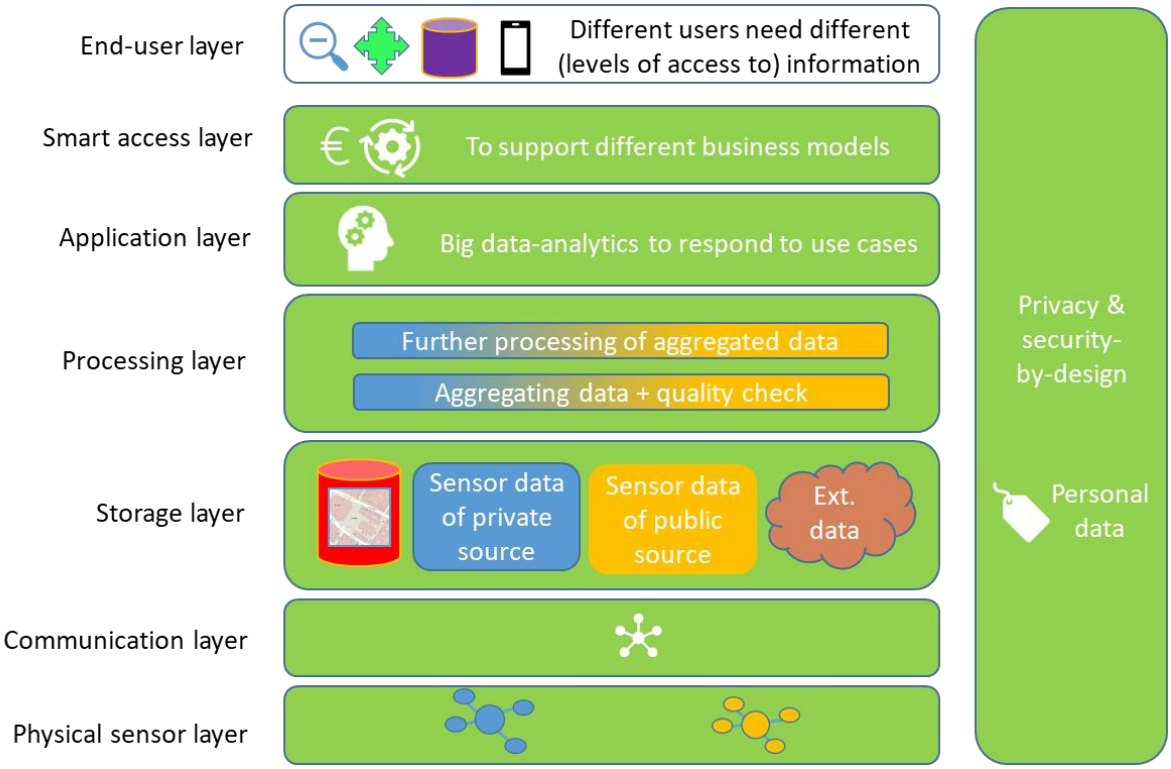

Figure 3: Data architecture of the desired solution

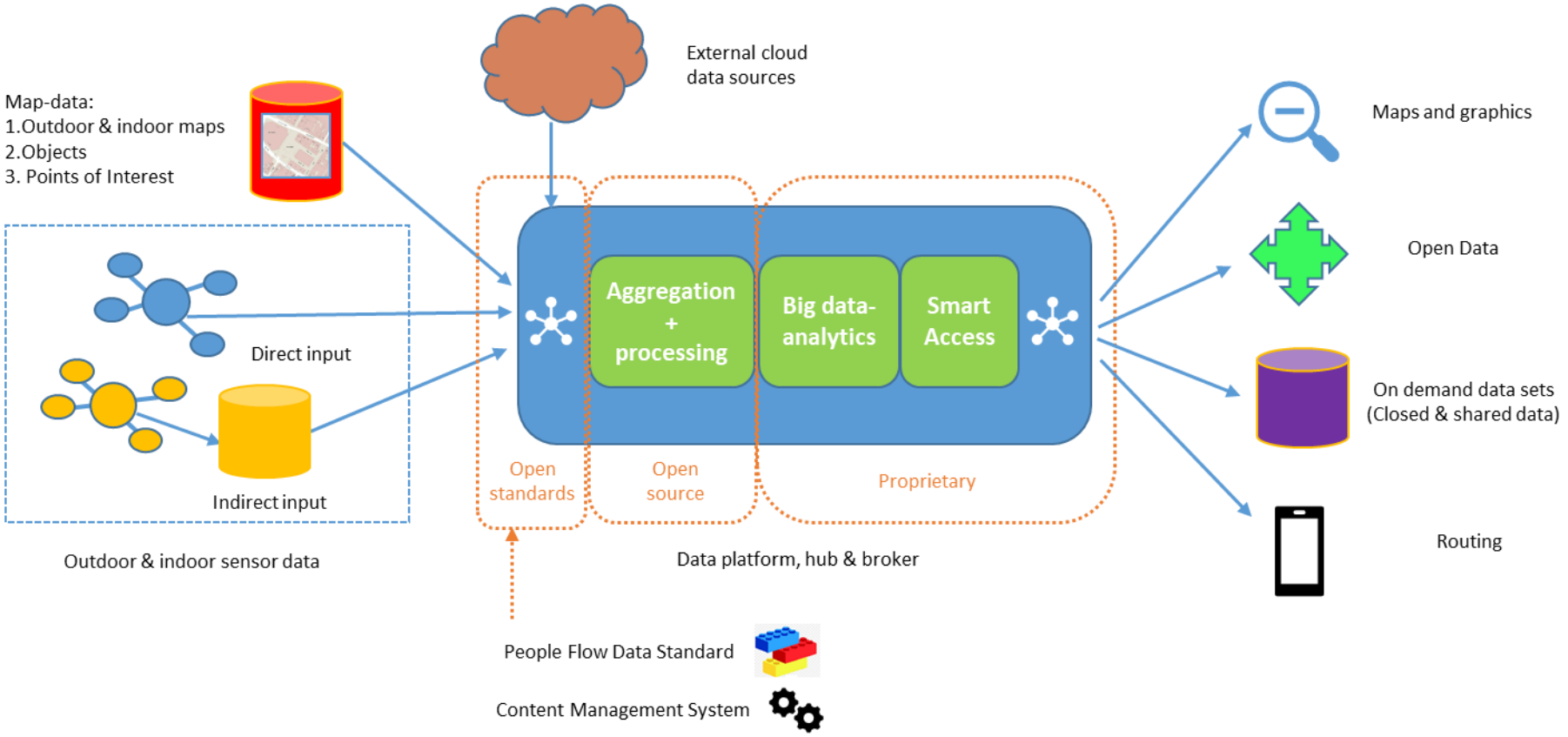

Figure 4: Visualisation of the desired solution