H2020-MG-2014-2015

MOBILITY FOR GROWTH 2014-2015

H2020-MG-2014_TwoStages

Sustainable Urban Consolidation CentrES for conStruction

Open Data accessibility and treatment document

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Dissemination level	*Public*
Workpackage and task	*WP2 / Task 2.5*
Author(s)	*LIST*
Contributor(s)	*En&Tech*
Due date of deliverable	31/10/2016
Submission date of deliverable	31/10/2016
Status (F: final, D: draft)	F

Document Control Sheet

Project number	**633338**		
Project Acronym	**SUCCESS**		
Work Package	**2**		
Version	1.0		
Issue	**Version**	**Dates**	**Content**
	0.1	07/12/2015	Initial version
	0.2	20/09/2016	Integration of quality measures for As-Is analysis and quality analysis for data sampling
	0.3	07/10/2016	Integration of open data survey results and conclusion. Integration of T. Schwartz comments.
	1.0	21/10/2016	Integration of partners reviews (S. Novellani) and external reviews (S. Turki) in final draft version

Classification of this report:

Draft	
Final	**X**
Confidential	
Restricted	
Public	**X**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

LIST project coordinator:

5 Avenue Des Hauts Fourneaux, Esch Sur Alzette 4362, Luxembourg

Contact details:

FERRERO Francesco, +352 275 888 22 27, francesco.ferrero@list.lu
GUERLAIN Cindy, +352 275 888 67 46, cindy.guerlain@list.lu
SCHWARTZ Thomas, +352 275 888 25 21, thomas.schwartz@list.lu

Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

1 Executive Summary

In the task 2.5, we set up principles for measuring the quality of the main dataset collected (deliveries and pickups) and giving feedback to data collectors on pilot sites. Focused on the purpose of the KPI computation for the As-Is analysis (task 2.4), 46 controls have been designed based on 3 main data quality criteria: completeness, consistency and accuracy. These controls have been run at each release of the data collection and data quality reports were issued per pilot.

The data quality activities have helped the pilots to increase the level of quality of the data collected over a long period of collection.

The majority of the data required for KPI computation are fit for their intended purpose (with a score above 75%) although a minor set of data have been more difficult to collect. For 3 pilot sites, the KPI computation makes use of at least 60% of the data collected. For the remaining pilot, the one that has collected the more data on deliveries and pickups than the three other pilots, two KPs make use of 30% to 50% of data collected while the other KPIs are at the level of the other pilots.

One can therefore be confident in the representativeness of the KPI computation for each pilot site and then in the representativeness of the As-Is analysis.

We have analysed the quality of the data for the purpose of sampling for simulation, in order to enlighten the decisions to be taken in the subsequent work packages and activities of the project. We notice that the data are generally of a sufficient quality for the purpose of simulation, but some points still have to be taken into consideration:

- Some data (such as names of suppliers, of materials or addresses) may need some correction and curation before being completely suited for simulation.
- In some cases suppliers may use several different addresses to deliver or pickup materials. This varies among the pilots.
- The materials that are the most delivered or picked-up are different between the pilot sites.
  - For a same pilot, there might be major differences in the list of materials depending on the criteria used to determine the "most delivered/pickups" (either weight transported or frequency of pickups/deliveries).
- The factor that limits the load of the transport vehicle is mainly the volume of the material transported rather than the weight. This varies among the pilots and is more identified for pickups than for deliveries.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

The quality level measured on the main data collected gives confidence in the interest to release the data as open-data for the purpose of further research. Pilots were asked their position towards the release of their data as open-data. It appears that pilots and data owners are not ready to release data in their current state. Pilots have identified business and confidentiality risks in case of data release. Actions to reduce these risks need to be taken by the project before considering the release of the data.

Finally, in case of data release we propose the project to consider a dual licensing on the data (commercial vs. non-commercial), based on the creative commons licenses.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

2 Introduction

SUCCESS has chosen to target the construction industry as major impacting sector on city logistics which has un-exploited potentials of improvement of the efficiency of goods, waste and service trips in EU cities, by answering the challenges pinpointed by the European Commission and in particular by improving urban freight understanding and by introducing more resource-efficient, more environmental-friendly, safer and seamless supply chain innovations.

The ***D2.5 Open Data accessibility and treatment document*** is part of the quantitative analysis, measures and data collection work package of the SUCCESS project (WP2) which maps and evaluates the current situation (AS IS) and evaluates it according a set of common Key Performance Indicators (KPIs).

This deliverable aims to determine the capacity of the project to release data collected in the work package as open data and the conditions to release the data. In order to determine this capacity, we have to consider two points of view: the quality of the data collected on the one hand and the benefits and risks to release such data for the data owners, the data collector and their stakeholders on the other hand. The quality of the data is determined in the light of the purpose of the data collection. In the second work package of the SUCCESS project, the purpose of the data collection is multifaceted: the data collected by the construction pilot sites is intended to be used primarily for the As-Is analysis (task 2.4 and D2.4) but also for the sites design solution (task 4.2).

This deliverable is structured as such: the principles and main results to ensure the quality of the data for the As-Is analysis are detailed in the first section. Then the quality of the data is analysed with regards to its usage as samples for simulation in subsequent tasks of the project. Next the benefits and risks associated to the release of the data are analysed. Finally, recommendations are made for releasing open data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

3 Principles of data quality for the As-Is analysis

In order to ensure the validity of the As-Is analysis performed in the work package WP2, we implemented principles of data quality. We identified the quality criteria, we implemented relevant controls and we set-up a quality management process in relation with other tasks of the work package (database structure definition and data management T2.3, data collection and analysis T2.4).

3.1 Data quality criteria

The main quality criterion for data is their fitness for purpose (in other words: are the data usable for their intended usage?).

The main purpose to collect data in work package 2 is to perform the As-Is analysis of the pilots by computing KPI. Therefore we started by mapping the categories of data to collect with the KPI defined in the task T2.2.

Then, still based on the generic definitions of data quality for open data[1], we identified the criteria contributing to the fitness for purpose. Such criteria are: data completeness, data consistency and data accuracy as represented on Figure 1.

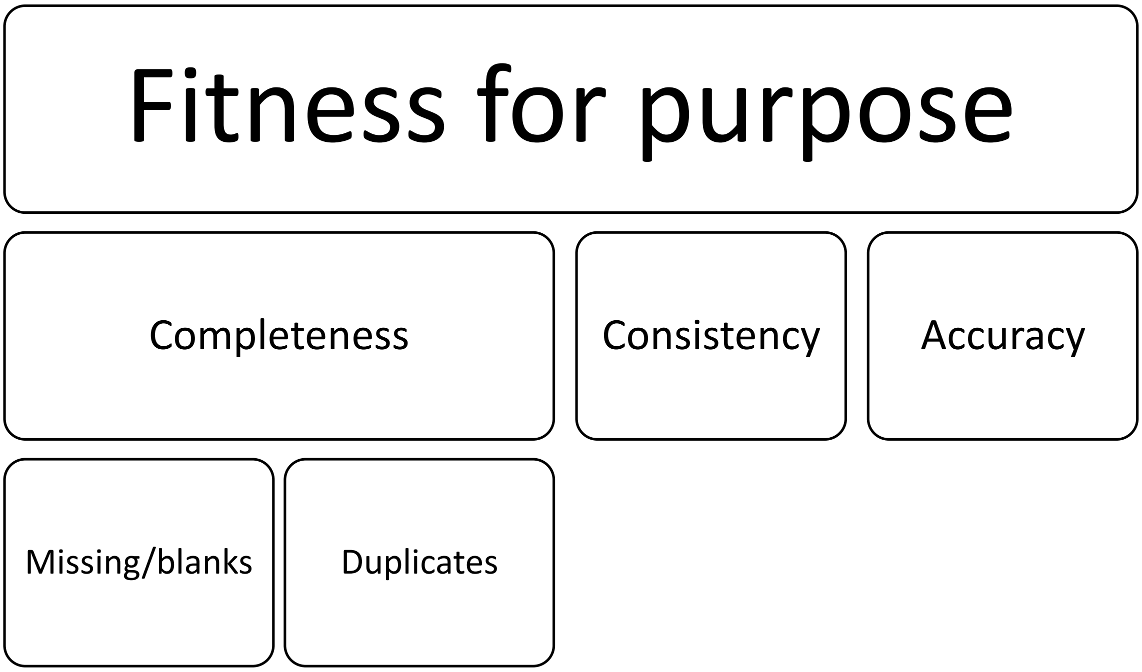


Figure 1: Overview of data quality criteria

[1] http://opendatahandbook.org/glossary/en/terms/data-quality/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

3.1.1 Data completeness

Data completeness criterion consists in determining for each data or data set, the range of acceptable blanks or missing data. The completeness criterion basically answers the question: *is the required data present or absent from the data-set?*

3.1.2 Data consistency

Data consistency criterion consists in determining if the data is consistent with other related data. Consistency is usually determined by cross checking the values of atomic data. Examples of consistency checks are: crossing time/distance/speed values according to the speed formula, crossing dates and times sequence according to the standard business process (e.g. for a delivery: loading < arrival in urban area ≤ arrival near site ≤ arrival at site ≤ unloading start < unloading end…)

3.1.3 Data accuracy

Data accuracy criterion consists in determining if the level of accuracy for the data collected is in line with the intended purpose of the data (in our case the computation of the KPI). By example we check if a recorded time is expressed in hours and minutes (seconds are over accurate), if the transported weight is expressed in the correct unit (kg), if recorded address can allow a precise location (i.e. we have at least name of the city and the country).

3.2 Data quality controls

Based on the mapping between the data and the KPI, and using the criteria identified beforehand, we determined controls that could be performed on the data.

Controls for **completeness** criteria consist in checking whether a value is input. 22 completeness controls were designed: one for each data directly related to a KPI computation.

Controls for **consistency** criteria were designed having in mind the type of the data and the KPI to compute and KPI computation formula. For example the collected data '*arrival time at construction site*' is a time data used to compute three KPIs: *congestion on construction site, truck waiting time outside the site, truck waiting time inside the site*. For each KPI the consistency control of the data was designed by cross checking the dependant data (by example the correct time sequence). In total 18 consistency controls were designed.

Controls for **accuracy** were scarcely ever designed due to the fact that the data type and precision was set in the database structure (see D2.3). Thus the control of data type and precision was performed by the database at the data import step. However we designed 6 accuracy controls, specifically for

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

addresses related data, by setting the accuracy level of an address to be properly identified by a geocoder. Then the control was: *does a geocoder returns a geocode (latitude, longitude) when submitting the address*?

The complete list of the 46 controls operated at data level is presented in *Appendix 1 – Implemented data quality controls.*

We then completed the mapping started in the previous activity to arrive to a complete view of controls performed on data to ensure their suitability for KPI computation. An excerpt of such mapping is presented in *Figure 2: Example of mapping between data, purpose (KPI), quality criteria* and controls.

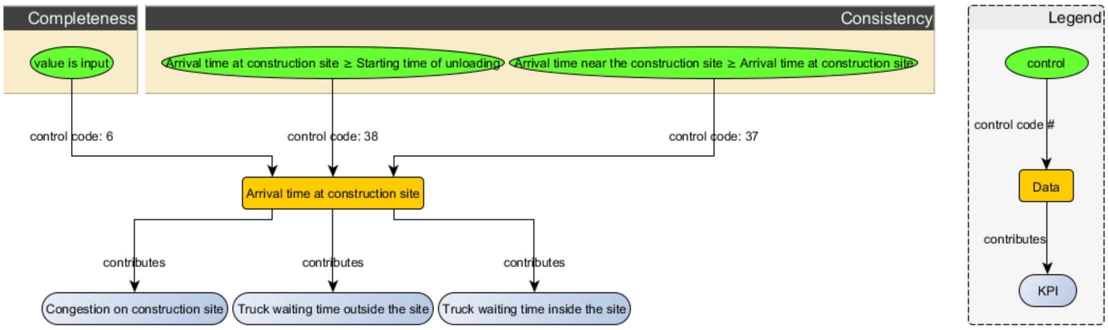


Figure 2: Example of mapping between data, purpose (KPI), quality criteria and controls

The complete mapping is presented in *Appendix 2 – KPI Data Controls Mapping (full view)*. This mapping allowed relating a control failure both to

- the data on which the control was performed,
- and to the purpose of the data, i.e. the KPI to be computed.

We were then able to report failed controls on each data of each pilot's dataset and provide an overview such as *Figure 3: Deliveries and pickups data quality* summary. We were also able to report the impact of the failed quality controls on KPI computation and provide an overview such as *Figure 4: Deliveries and pickups data fitness for purpose at the end of data collection*.

3.3 Data quality in the data management process

Criteria, levels and assessment methods are fixed to assess the fitness of the data for such purposes. The collected data are assessed following the applicable criteria and controls. Inexpediency detected between data and the intended purpose(s) is reported to the relevant partners (data collectors, task leaders,…).

Pilots' data quality is managed based on the data collected in task T2.4. Pilots' data are loaded by T2.3 leader (En&Tech) in the reference database. The data quality is controlled on the data loaded in the reference database. At each update of the reference database, T2.5 leader (LIST) runs the 46 quality controls through specific queries on every line recorded in the database.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

T2.5 leader (LIST) generates a data quality report. The report is specific to each pilot and is composed of three parts:

1. An overview of **data fitness for purpose** indicating for each data required for KPI computation the rate of fitness for purpose, and the distribution of errors among criteria completeness, consistency.
2. An overview of **KPI computation capability** indicating the number of data suitable for computation (i.e. without errors) with regards to the total number of data recorded (for the pilot).
3. A list of data correction suggestions.

T2.5 leader (LIST) issues data quality reports at each update of the database to T2.4 leader (VCF) and addresses a copy to T2.3 leader (En&Tech) for information.

T2.4 leader (VCF) takes the decision to request change or data correction to pilot coordinator (LIST, En&Tech, VCF, VPF) which in turn asks correction to pilot (TRALUX, CMB, VCF, Pavasal). If so a copy of the request is addressed to T2.3 leader for information. Data correction is performed only by the pilot's data collector.

3.4 Quality summary

At the date of September 22nd 2016, at least 5 updates of the reference database have been released since March 2016. At major releases, quality reports have been issued to pilots. The quality reports helped them to know where to improve the quality of their data so that their data can be used for the As-Is analysis. As mentioned in deliverable D2.4 we noted a significant improvement of the quality of the data during the data collection period.

As represented on *Figure 3: Deliveries and pickups data quality summary at the end of data collection*, we can note that, except three data that are between 50% and 66% percent of fitness for purpose (Arrival time in urban area, arrival time near the construction site and come back empty or loaded), the remaining of the key data required for the As-Is analysis are at least 75% fit for purpose.

At the end of data collection, the main type of error is incompleteness. It often means that required data were not collected for the delivery / pickup. This type of error gets harder to correct as the date of collection is far in the past. Then this type of error cannot be further corrected.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Figure 3: Deliveries and pickups data quality summary at the end of data collection

With regards to the capability to rely on the data to perform the As-Is analysis, one can note from *Figure 4: Deliveries and pickups data fitness for purpose at the end of data collection* that the KPI computation makes use of at least 50% of the data collected except for two KPIs for one pilot site. However this Pilot is the one that has collected the most deliveries and pickups data, therefore the number of data fir for purpose in absolute value is at the same level of higher than other pilots. This gives confidence in the representativeness of the KPI for each pilot site and subsequently in the representativeness of the As-Is analysis for each pilot site.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

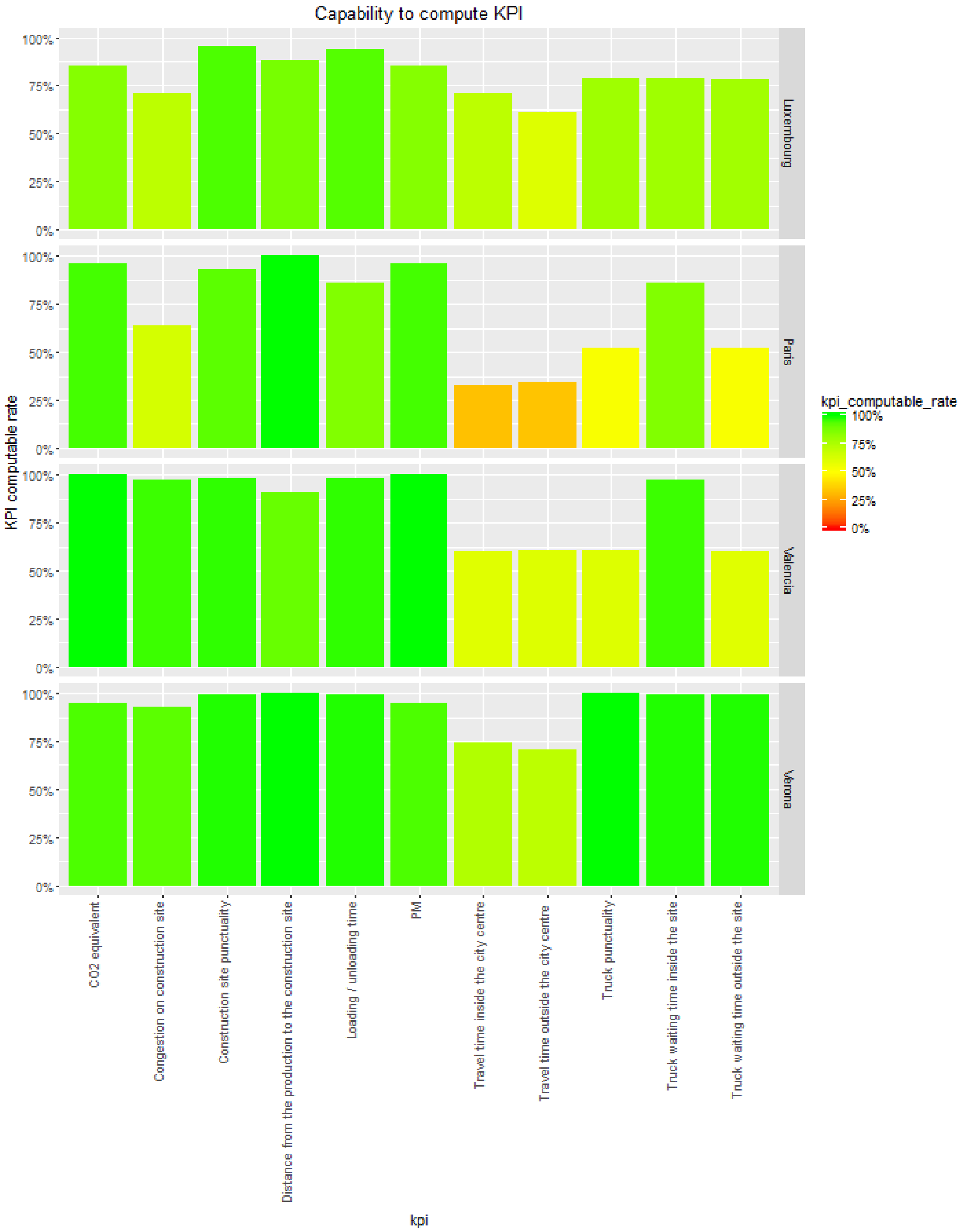


Figure 4: Deliveries and pickups data fitness for purpose at the end of data collection

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

4 Quality analysis of data sampling for simulation

Apart from quality for the purpose of As-Is KPI computation described in the previous section, we performed further data quality analysis on the whole data set in the light of a potential use as samples for simulation.

Part of the data required by the algorithms developed in T3.1 to perform the simulation in task T4.2 were directly required for computation of the KPI for the as-is situation in task T2.4. For these data, data quality controls have been performed regularly during the data collection to ensure a minimal level quality and reliability of the KPI computation. However, for the data required for simulation that were not directly required for computation of the as-is KPI, such data quality controls have not been performed.

Then the following subsections detail additional data quality analysis so as to provide insights on the level of quality for the purpose of simulation.

4.1 Deliveries data

4.1.1 Suppliers

4.1.1.1 Completeness

We check the number of times the supplier name is input per construction site deliveries. The lack of a supplier name implies the impossibility to identify the supplier of the delivered material.

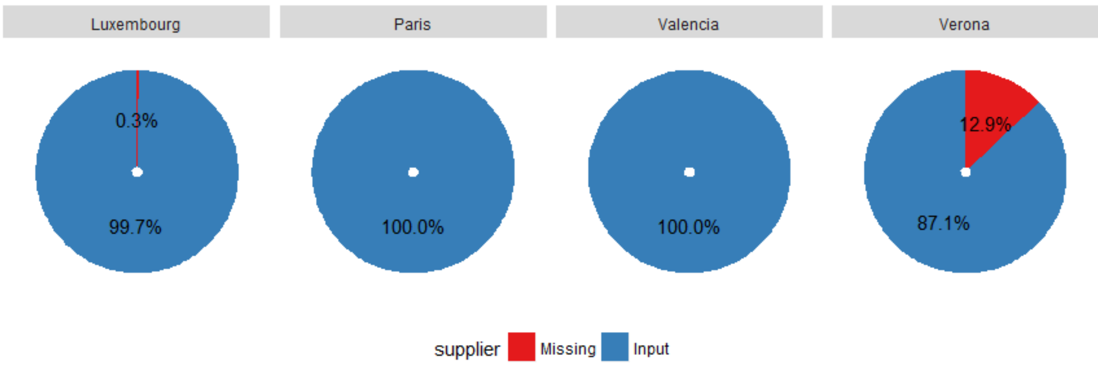


Figure 5: completeness of supplier's name per pilot site

We notice that **for 3 pilot sites, the supplier name can be used to identify precisely the supplier. For the last site, the rate of incompleteness is quite higher and more than 10% of the deliveries cannot be linked to one supplier**.

4.1.1.2 Consistency

We determine the consistency of suppliers' names by counting the number of distinct suppliers names related to a same address.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

When this count is > 1 there is a high probability that a supplier has been described in a non-consistent manner (i.e. in a non-unique way).

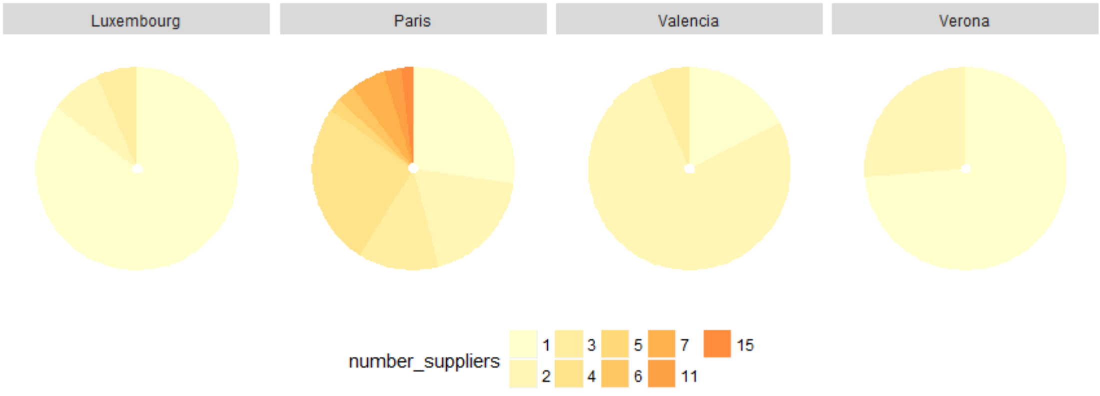


Figure 6: Number of distinct supplier's names for a unique address per pilot site for deliveries

We notice that the situation is different for the pilots. For two sites, the supplier name is written quite consistently with 75% of loading location addresses counting one and only one supplier, while the remaining 25% of loading location addresses counts less than 4 names for a unique address. For a third site, the situation is quite similar but the thresholds are a bit higher (20% of addresses count 1 supplier, 75% of addresses count 2 suppliers). For the last site, the situation is very different since only 25% of addresses count one and only one supplier, but more than 33% count 4 or more suppliers' names. This indicates that for this site, **data curation on suppliers name will be necessary to correctly identify suppliers' loading location addresses**.

4.1.2 Loading location

4.1.2.1 Consistency

We determine the consistency of suppliers' addresses by counting the number of distinct loading location(s) address(s) per supplier.

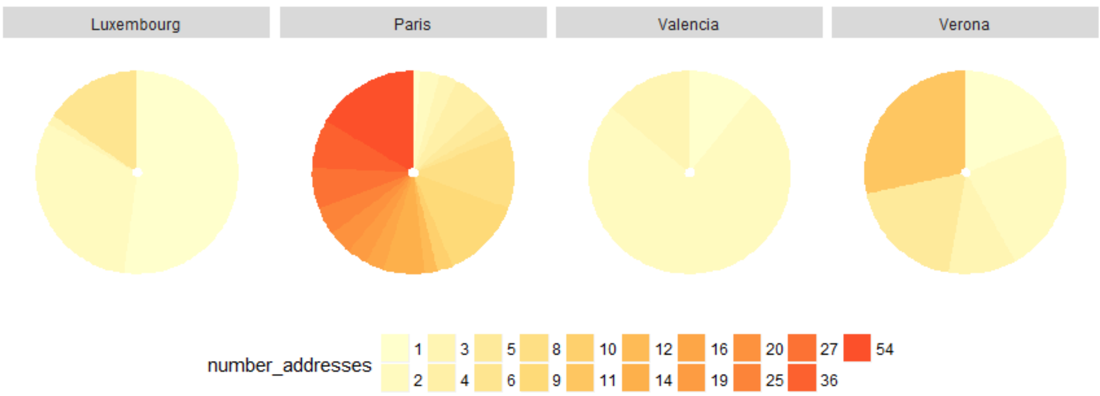


Figure 7: Number of distinct addresses for a unique supplier, per pilot site

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

We notice that there is no general rule that allows relating one and only one loading location address to one supplier. The fact of identifying more than one loading location address to a unique supplier might be explained by - an inconsistency in the writing of the loading location address, - or multiple addresses corresponding to various warehouses or production sites for the supplier, - or a combination of both.

In general there is a 1-to-many relationship between the supplier and the address of loading location which has to be taken into account for simulation. Then before using the data for simulation **it should be clarified with the construction site what part of the material comes from which depot and what part comes directly from the production site.**

4.1.2.2 Accuracy

In order to determine accuracy of suppliers' loading location addresses per pilot site, we counted the number of loading location addresses that were usable for computing distances (i.e. that could be geocoded) and compared it to the total number of suppliers' loading location addresses per pilot site.

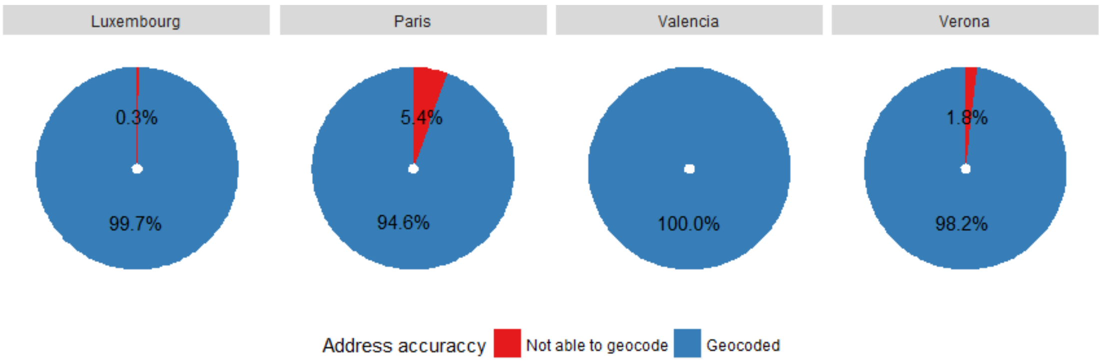


Figure 8: Rate of geocoded loading location addresses per site

We see from this analysis that **inaccurate loading locations are quite scarce for most construction sites, with inaccuracy less than 6% for all sites**. This indication shows that **loading locations addresses recorded in the delivery tracking board are reliable enough for geocomputation**.

4.1.3 Material

4.1.3.1 Completeness

We check the number of times the material name is input per construction site deliveries.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

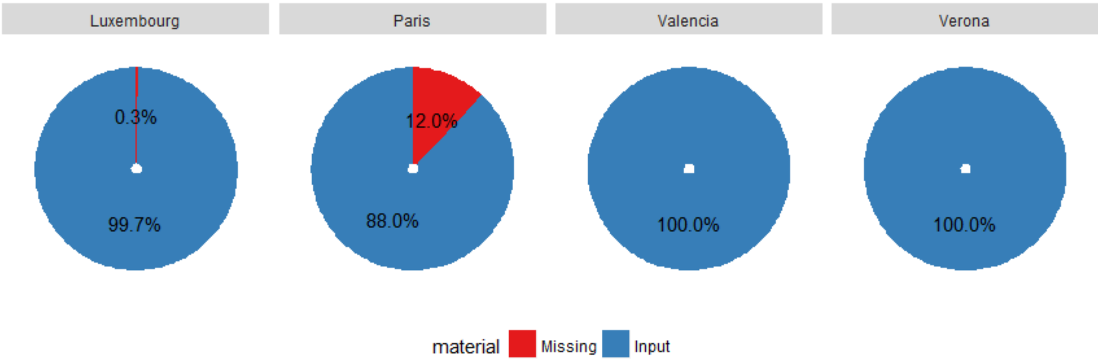


Figure 9: Completeness of material name per pilot site

We notice that **for 3 pilot sites, the material name can be used to identify the material delivered**. **For the last site**, the rate of incompleteness is quite higher and **more than 10% of the deliveries do not specify the type of material delivered**.

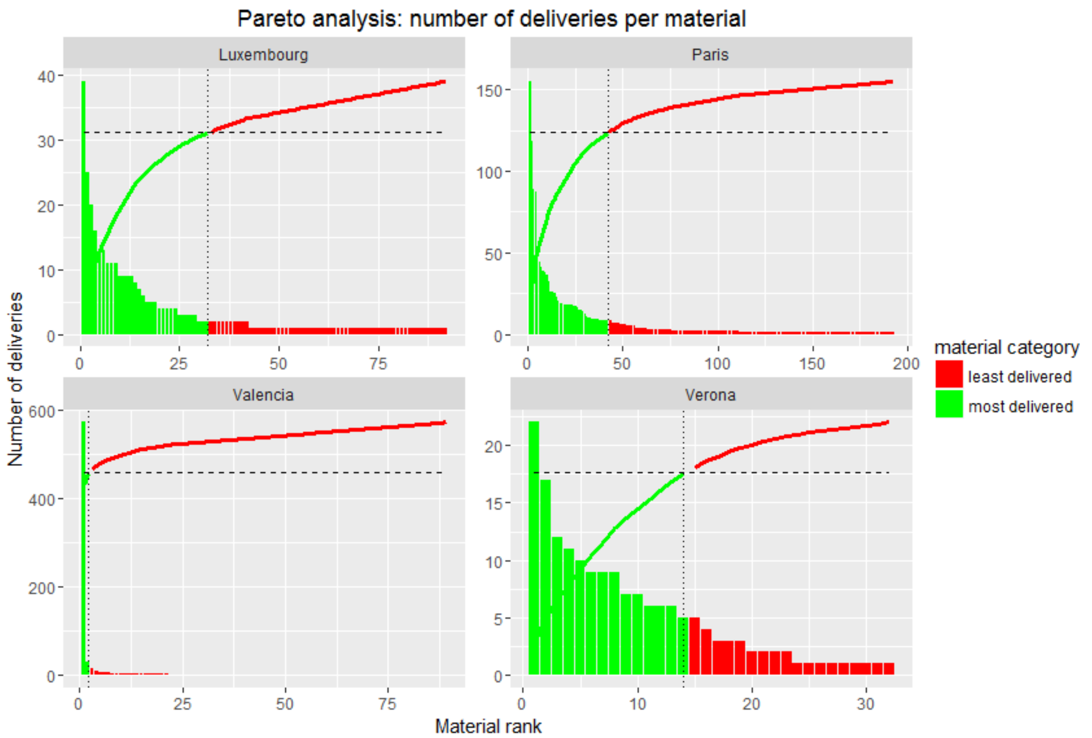
4.1.3.2 *Materials most delivered*

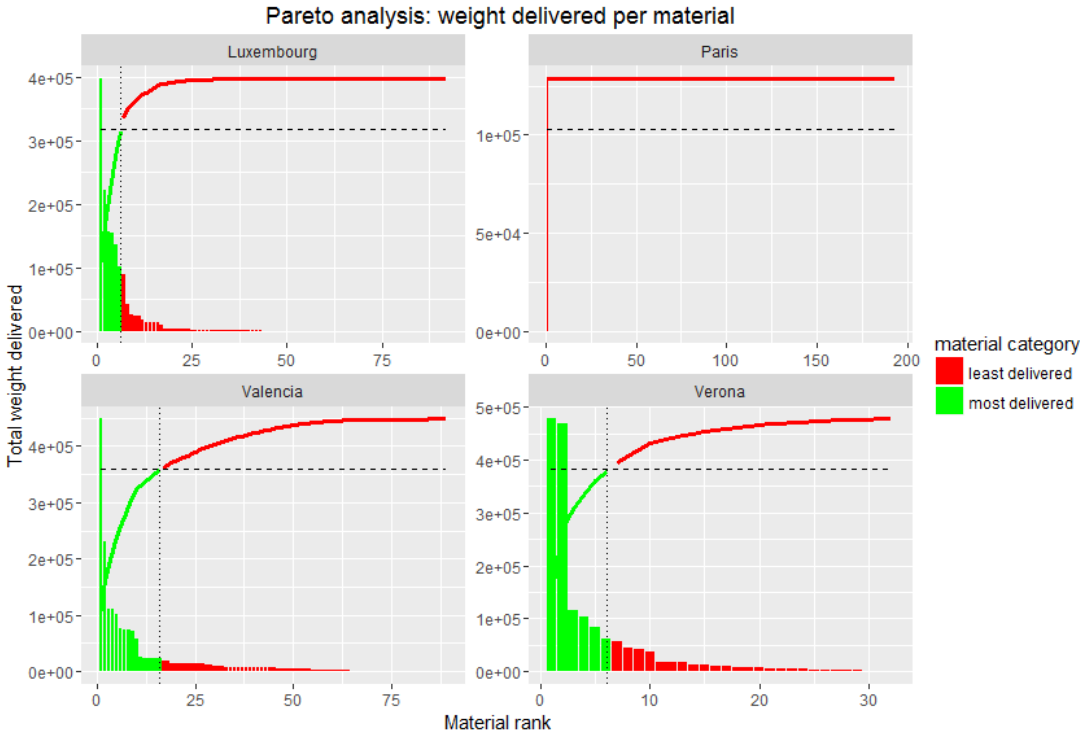
The choice of the representative materials to perform the simulation of the to-be for each pilot site is critical. In order to identify such representative materials per construction site, we determine the materials that are the most delivered in weight delivered and in number of deliveries (frequency).

The two analyses for each criterion are distinct, but the method is common. We represent the total for each material delivered on the site (bars on y axis) and we order each material from the most delivered to the least one (Material rank on x axis). We also represent the cumulative sum for all deliveries on the site (line). Last we represent the 80% threshold of the delivery criteria (weight or frequency) as the horizontal dashed line and the number of distinct materials under the threshold as the vertical line at the cut between the cumulative sum and the 80% threshold. All materials ranked before the cut-off (in green) represent the most delivered materials. All material ranked after the cut-off (in red) represent the least delivered materials. For reading convenience, the names of materials have not been reported on the two graphs, but the complete data table is reported in *Appendix 3 – List of most delivered and collected materials* per site.

For the purpose of simulation, this analysis provides an indication on which material deliveries are most representative for each site depending on the criteria used to determine the "most delivered" (total weight transported or total number of deliveries).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.





The figure shows that **the situation is very different according to the site and to the criteria used to determine the most delivered materials (weight or frequency)**. For 2 out of 4 pilots sites the frequency criteria implies more distinct materials than the weight criteria, while it is the opposite for one site and cannot be computed due to lack of data for the last site.

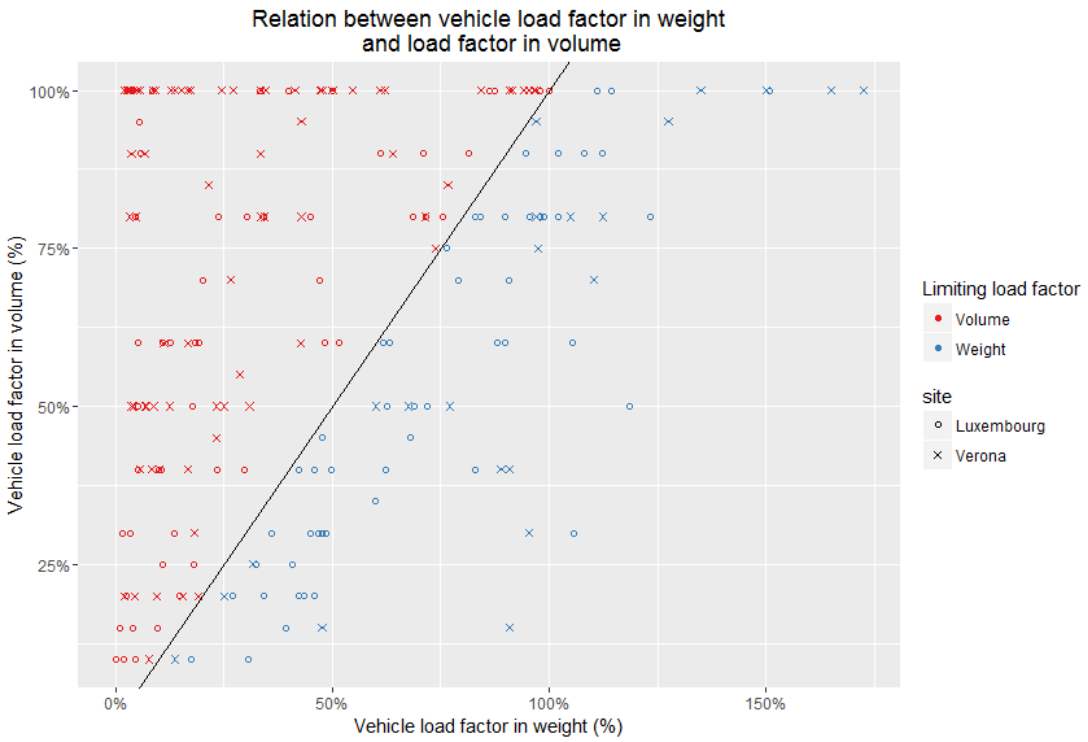
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

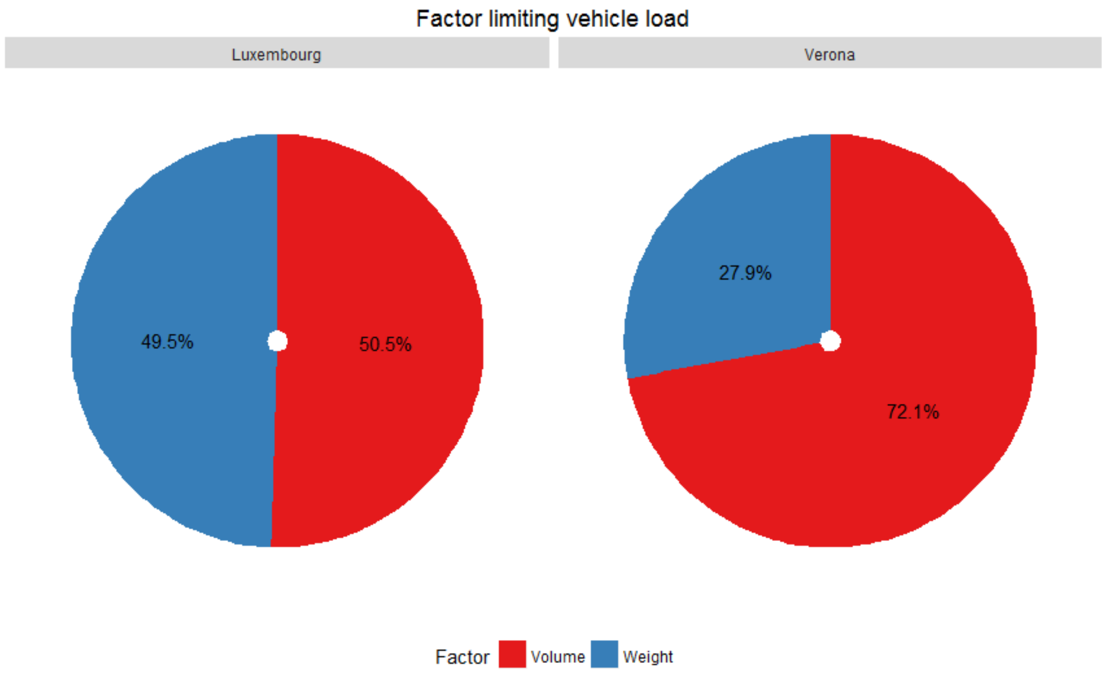
It might be necessary to perform the simulation by using materials corresponding to each criterion and to carefully examine the differences (or lack of difference) before drawing any conclusion on the simulation.

4.1.4 Vehicles capacity limits

Determining the limiting factor for the delivery vehicles capacity can provide insights for identifying representative hypotheses for the simulation. In order to determine this limiting factor, we first determine the weight capacity of the vehicle for each delivery. The weight capacity is the gross weight of the vehicle minus the net weight. Then we compute the load factor of the vehicle in weight as the ratio material weight / weight capacity. Last we relate this load factor in weight to the load factor in volume collected during the data collection. Note that due to the lack of data for two sites (Paris and Valencia) we were able to perform the analysis on only two pilot sites (Luxembourg and Verona).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.





The two figures show that **vehicles' load is limited by the volume capacity rather than the weight capacity** 3 times out of 4 for Verona, while for Luxembourg the two criteria are almost balanced. This fact has to be taken into account in the simulation for determining the allocation of vehicles.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

4.2 Reverse logistics (pickups) data

4.2.1 Suppliers

4.2.1.1 Completeness

We use the same method as mentioned in section 4.1.1.1, but applied to pickups data.

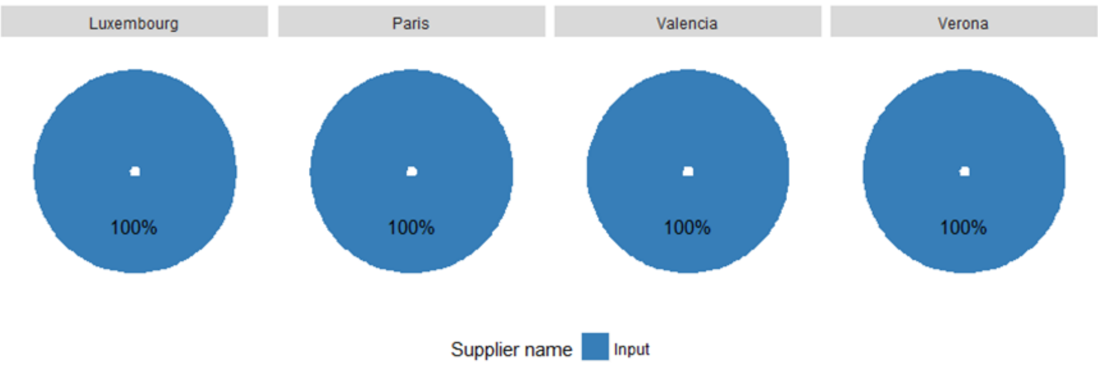


Figure 10: Completeness of supplier's name per pilot site for pickups

We note that all sites have **no completeness issue on reverse logistics suppliers**.

4.2.1.2 Consistency

We use the same method as mentioned in section 4.1.1.2, but applied to pickups data.

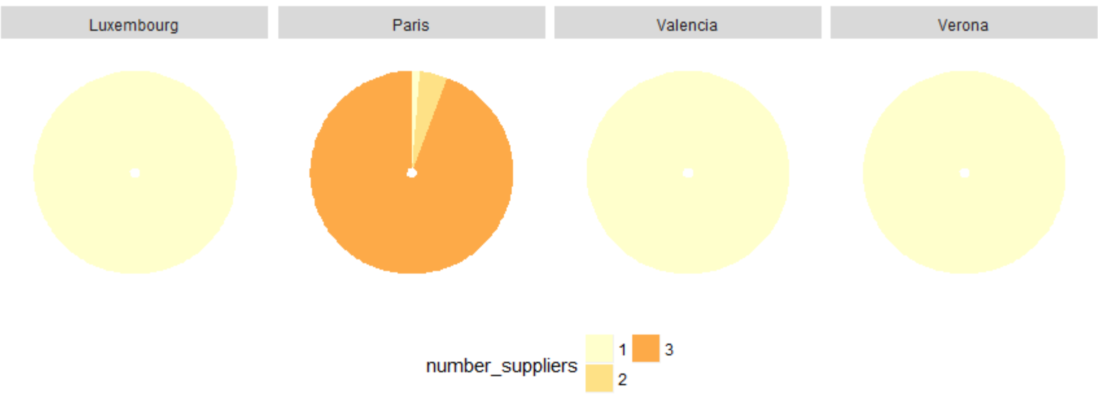


Figure 11: Number of distinct supplier's names for a unique address per pilot site for pickups

We note that **3 out of 4 pilots have no consistency issue on suppliers address**. For the remaining site most loading location addresses count 3 different

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

suppliers' names. This indicates that **for this last site, data curation on suppliers name will be necessary to correctly identify suppliers' unloading location addresses**.

4.2.2 Unloading location

4.2.2.1 Consistency

We use the same method as mentioned in section 4.1.2.1, but applied to pickups data.

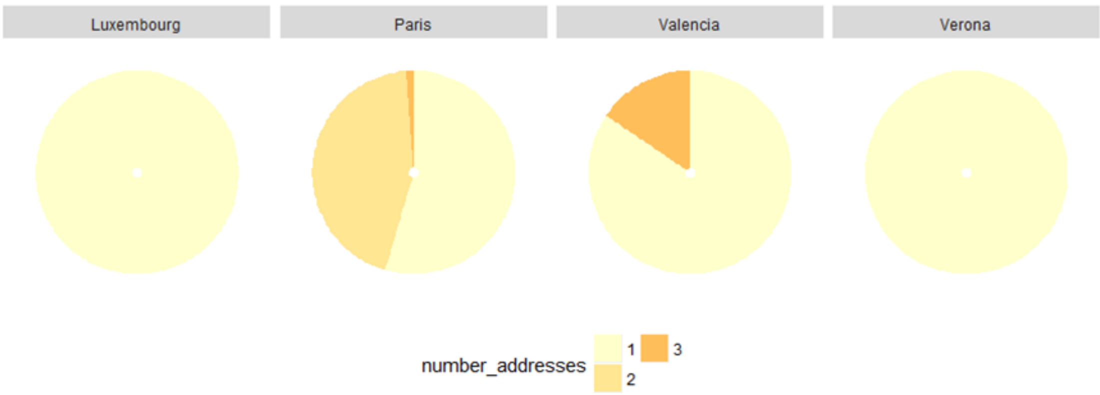


Figure 12: Number of distinct addresses for a unique supplier, per pilot site, for pickups

We note that half of the sites have no consistency issue on unloading locations. One site has recorded 3 distinct addresses for around 15% of reverse logistics suppliers, while the remaining site has recorded 2 distinct addresses for almost half of its reverse logistics suppliers.

For these **minor cases of 1-to-many relationship between the supplier and the address of loading location**, before using the data for simulation it **should be clarified with the construction site what part of the material goes to which recycling or waste facility**.

4.2.2.2 Accuracy

We use the same method as mentioned in section 4.1.2.2, but applied to pickups data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

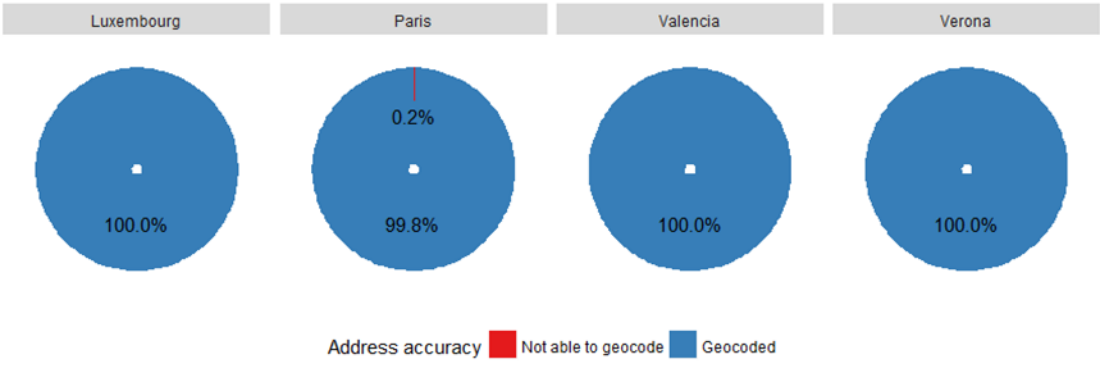


Figure 13: Rate of unloading location addresses geocoded per site

We note that we have almost no accuracy issues on addresses of unloading location. The error for only one site is not significant.

4.2.3 Material

4.2.3.1 Completeness

We use the same method as mentioned in section 4.1.3.1, but applied to pickups data.

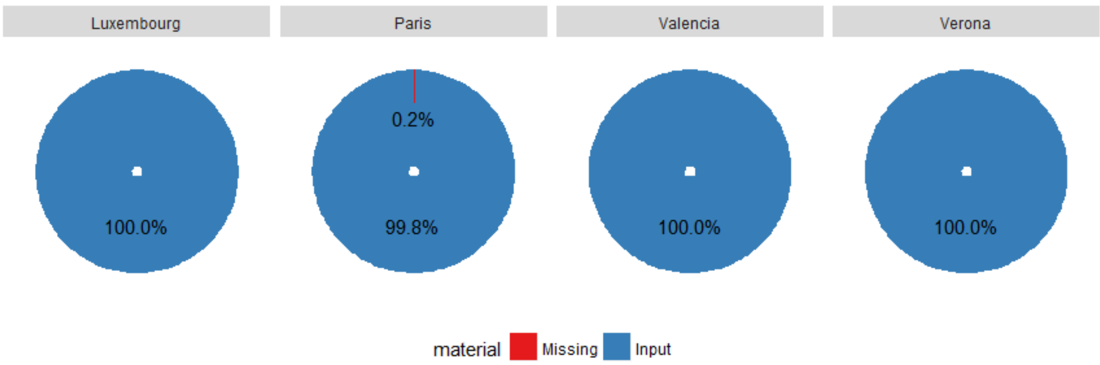


Figure 14: Completeness of material name per pilot site for pickups

We note that we have almost no completeness issues on material. The error for only one site is not significant.

4.2.3.2 Materials most collected

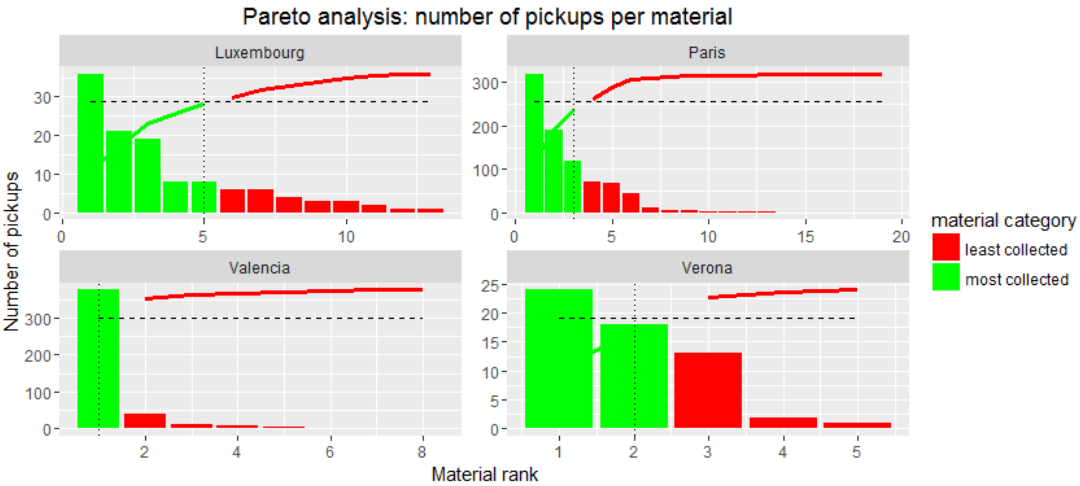
The choice of the representative materials to perform the simulation of the to-be for each pilot site is critical. In order to identify such representative collected

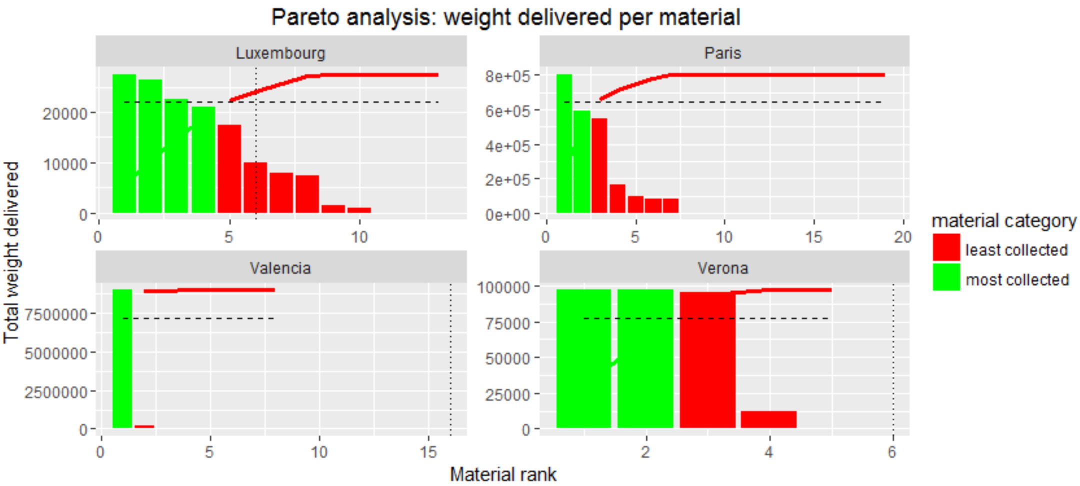
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

materials per construction site, we determine the materials that are the most collected in weight collected and in number of pickups (frequency).

The two analyses for each criterion are distinct, but the method is common. We represent the total for each material collected on the site (bars on y axis) and we order each material from the most collected to the least one (Material rank on x axis). We also represent the cumulative sum for all pickups on the site (line). Last we represent the 80% threshold of the pickup criteria (weight or frequency) as the horizontal dashed line and the number of distinct materials under the threshold as the vertical line at the cut between the cumulative sum and the 80% threshold. All materials ranked before the cut-off (in green) represent the most collected materials. All material ranked after the cut-off (in red) represent the least collected materials. For reading convenience, the names of materials have not been reported on the two graphs, but the complete data table is reported in appendix.

This provides an indication on which material pickups are most representative depending on the criteria (total weight transported or total number of pickups) for performing simulation for each site.





This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

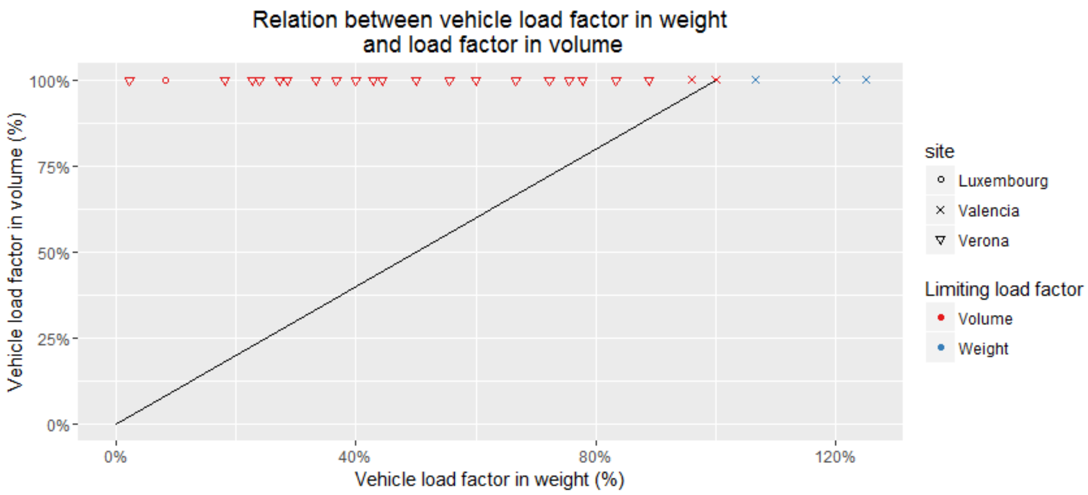
We note first that there are fewer categories of collected materials than categories of delivered materials which is quite relevant.

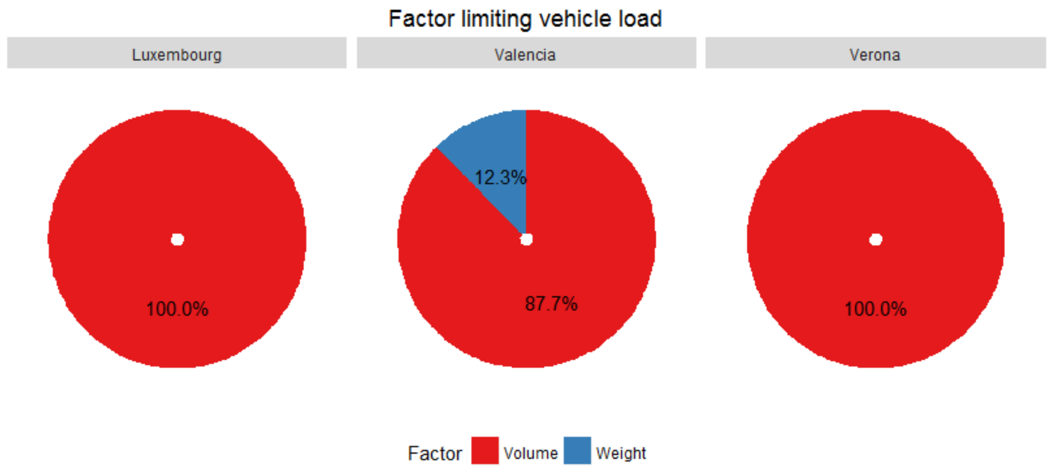
We also note that like for deliveries, the situation is different from site to site. One site has a balance in the criteria used to determine the most delivered materials (weight or frequency). For one site, no weight data were collected for pickups so we could only list the most collected materials in frequency. For the two remaining sites, the weight criterion is the dominant factor to determine the most delivered materials.

It might be necessary to perform the simulation by using materials corresponding to each criterion and to carefully examine the differences (or lack of difference) before drawing any conclusion on the simulation.

4.2.4 Vehicles capacity limits

We use the same method as mentioned in section 4.1.4, but applied to pickups data.





This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

First it is worth to mention that very few complete data were available for this specific analysis, so the conclusion might not be representative of the whole population.

We note that the factor limiting the load of the reverse logistics vehicle is clearly the volume rather than the weight.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

5 Benefits and risks related to the data collection opening

Data collected in the work package 2 of the SUCCESS project are business data related to delivery, pickup, and activity monitoring. These data have been collected and are mostly owned by private companies. Therefore it seemed necessary to assess the potential benefits and risks, including business related risks, related to the opening of the data collection before deciding which data to open and under which conditions.

5.1 Motivation for opening collected data

Since its definition, the SUCCESS project claims that one of its objectives in the work package 2 is to evaluate the capability to use collected data for new academic research project. It is clear that taking into account the quality measured and analysed beforehand, the data collected in the framework of the project are of a great potential for further research, in particular on construction logistics optimisation and Construction Consolidation Centers modelling.

The generic benefits that the project, the pilots and the data owners can get from opening part of the collected data are multiple. The release of open data would show the contribution of the project and its consortium to the European Commission strategy to open data. The data owners (i.e. the pilots) would gain a positive impact on their organisation's reputation by demonstrating their transparency thus contributing to their societal responsibility. Next the project and its consortium would get feedback from the users of the released data who may provide original analyses[2] on the datasets. Additionally the project and its partners could get access to new services designed by users of the datasets, should the released data be used / included in such services. Last by setting an adequate licensing schema on the released data, the project and its consortium may benefits from commercial licences sold on the data.

However releasing the data require taking into account the risks raised by the pilots and the data collectors.

5.2 Major risks identified in relation to open data release

A survey has been designed to collect the potential risks related to the opening of their data. Questions were issued about information security risks (in particular confidentiality infringement) and business risks related to each category of pilot's data. The survey has been addressed to Pilots or Pilot coordinators. Pilots were asked to assess their perception of risk in general and per category of data collected, then their willingness to open the data collected as-is (without any modification) or after anonymisation of sensitive data. Last, Pilots were

[2] i.e. additional analyses to the ones performed in the SUCCESS project

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

asked their expectation towards intellectual property conditions for opening their data. We report hereafter the main conclusions of this survey.

In the survey pilots and data collectors were proposed to assess their exposition to risks related to the release of open-data with the following categories of risks and examples.

Information Security risks related to data opening are globally risks to divulgate confidential or private information:

• *Privacy divulgation*: Name or any information (age, gender, coordinates…) of a specific person. Note that this case is not applicable to the data collected in the T2.4 task of the SUCCESS project, since no personal data have been collected.

• *Confidential information divulgation*: divulgation of Pilot's classified information or information that have been explicitly mentioned as confidential to the Pilot by its stakeholders (suppliers, its subcontractors, its Customer…).

Business risks are globally risks to divulgate not confidential but business sensitive information:

- *Competition risks*: use of information by competitors to benchmark or retro-engineer business processes (e.g. number of deliveries per day, names of suppliers or hauliers, type of material delivered by week or month…)
- *Reputation risks:* misuse of information to threat the reputation of the company (e.g. low performance, delays in construction…)
- *Compliance risks:* use of information to detect and expose potential compliance failures (e.g. non respect of environmental standards for some vehicles, non-respect of regulations on driving time or working time…)

Examples of such business sensitive information could be:

- o a customer may consider business sensitive the list of material (or the amount of a specific material) delivered to construct its building
- o a supplier may consider business sensitive the name of the haulier he contracted to deliver its material
- o the pilot may consider business sensitive the cost of waste, of transport or treatment of unsorted bins

It is worth to mention that one out of the four pilots is not a direct partner of the project. Therefore the project partner coordinating this pilot has established a specific agreement with the construction company. This agreement sets out that the data collected is confidential to the project consortium and prevents any usage out of the project scope. Thus the release of this specific pilot's data is a priori excluded unless an agreement is found with the construction company.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

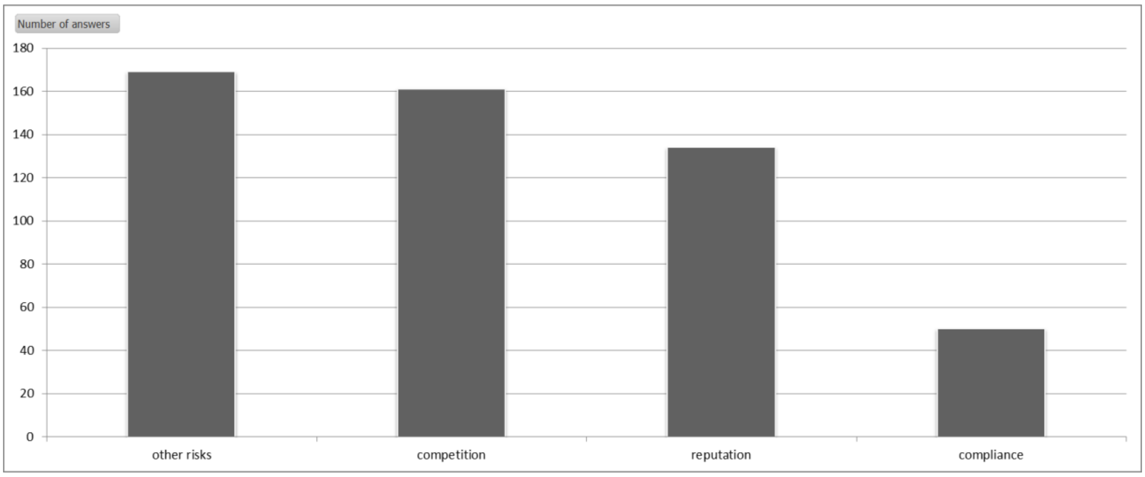


Figure 15: Major business risks related to release of data collection

From a global perspective, the **major business risks** identified by pilots and data collectors (see Figure 15) are *others* business risks (although the risks falling in the 'other' category is not further detailed by the respondents except in rare cases where *privacy* is mentioned), then *competition* risks, then *reputation* risks. In a lower extent *compliance* risks are mentioned.

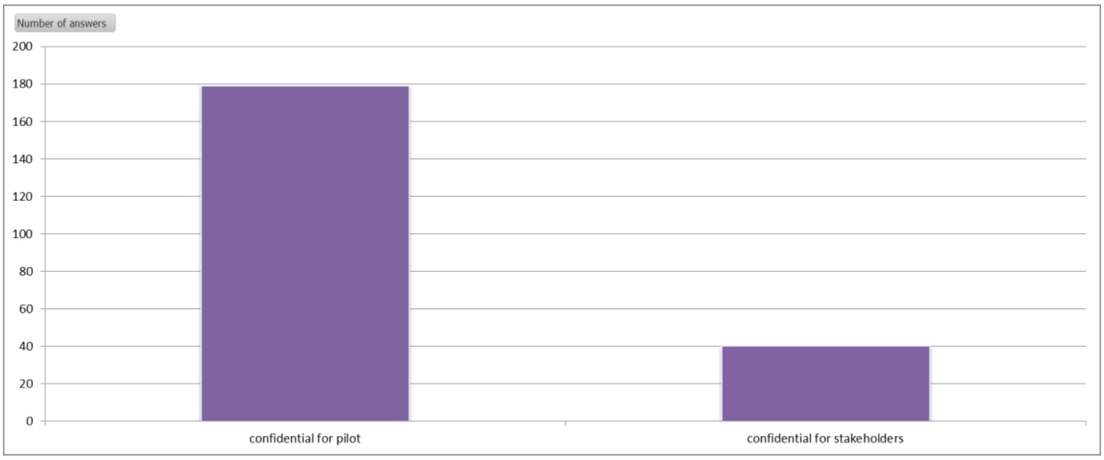


Figure 16: Actors for which the datasets are confidential

The **confidentiality sensitiveness** of the dataset (see Figure 16) is mostly related to the *pilot* itself and much more rarely related to the *pilot's stakeholders*.

The datasets that are the most subject to confidentiality (see Figure 17) are first *delivery and pickups*, then *logistic activities* followed by *material installation*. For the very large majority of dataset the confidentiality it mostly or totally related to the *pilot itself*, but notably for the *haulier route* dataset this confidentiality issue is balanced between *the pilot* and *its stakeholders*.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

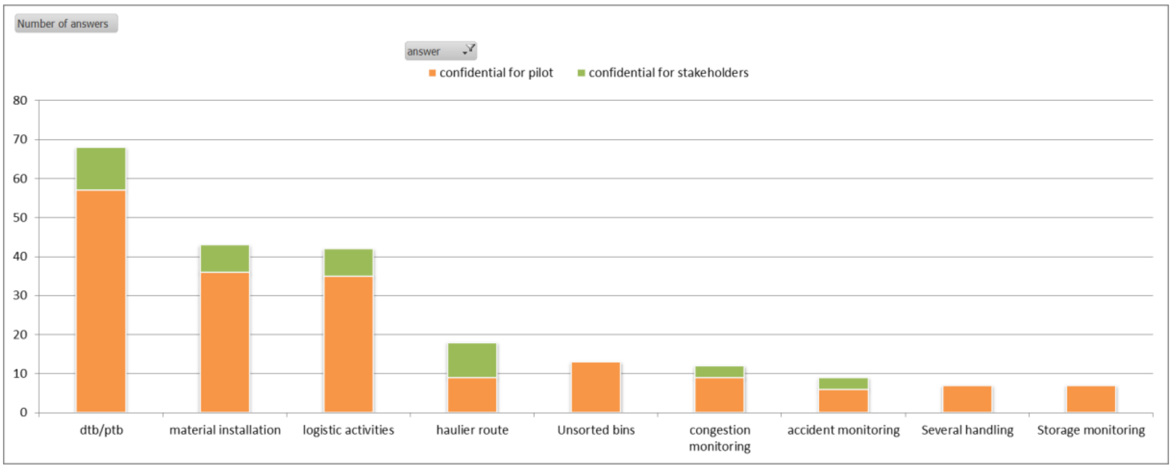


Figure 17: Datasets' sensitiveness to confidentiality

The datasets that are the most exposed to **competition** risks (see Figure 18) are first *delivery and pickups*, then *material installation* and *logistic activities*.

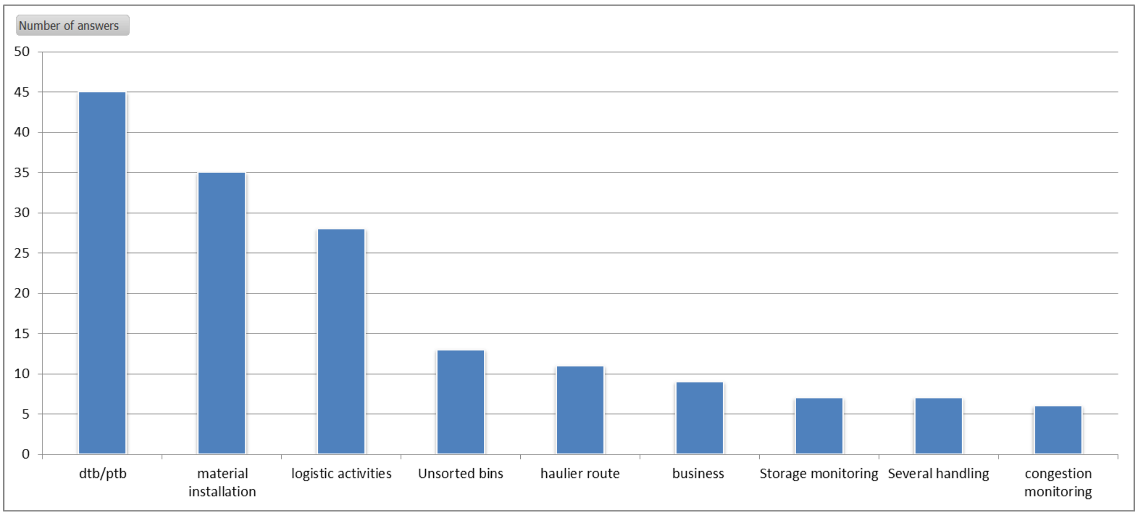


Figure 18: Datasets' exposition to competition business risk

The datasets that are the most exposed to **reputation** risks (see Figure 19) are the same: first delivery and pickups, then material installation, the remaining datasets being less exposed to such risk.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

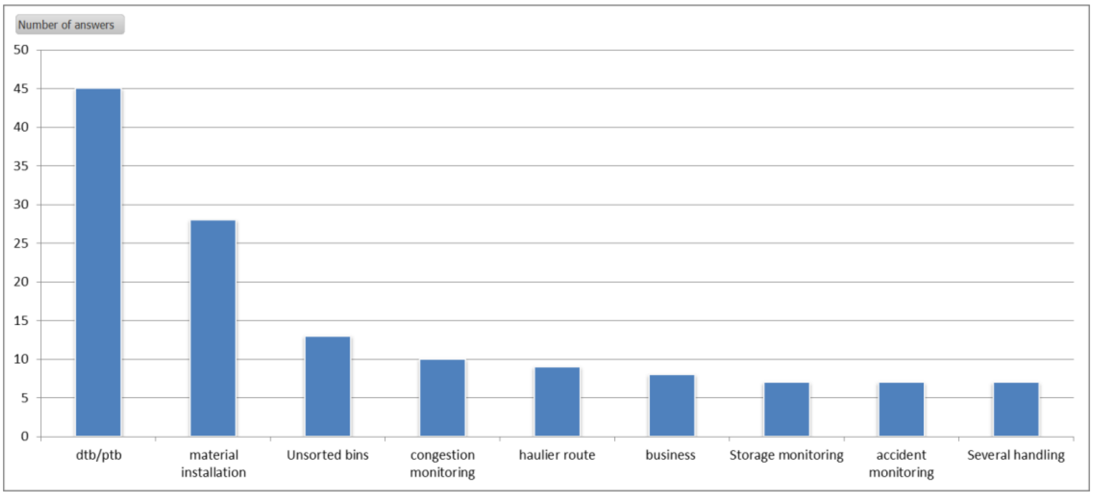


Figure 19: Datasets' exposition to reputation business risk

The datasets that are the most exposed to **compliance** risks (see Figure 20) are *material installation* and in a lower extent *congestion monitoring*, the remaining datasets being less mentioned as exposed to compliance risks.

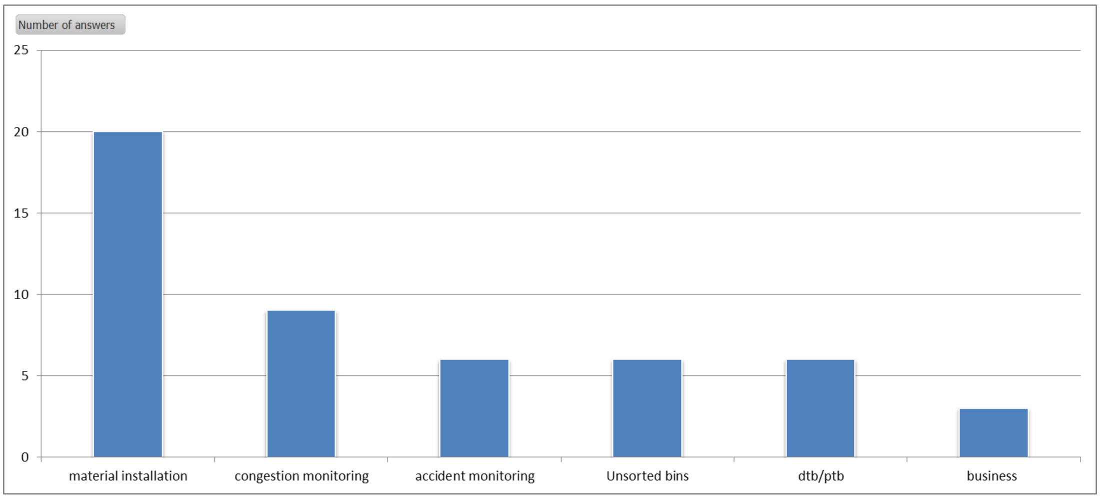


Figure 20: Datasets' exposition to compliance business risk

5.3 Main data with risks identified in relation to open data release

As a summary of the previous section, the datasets *deliveries and pickups*, and *material installation* are identified as both most exposed to business risks and subject to confidentiality considerations. Within this datasets the data that are the most exposed to risks (all categories of risks considered), are the name of the construction site, the activities dates and the name of the materials.

This is confirmed by the global consideration of the pilots on the confidentiality of their datasets (see Figure 21). Indeed, pilots mention to have a major confidentiality concern related to the *deliveries and pickups*, and *activity monitoring* datasets, while the confidentiality sensitiveness is lower for posteriori analysis but still remains a concern for the majority of the pilots.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

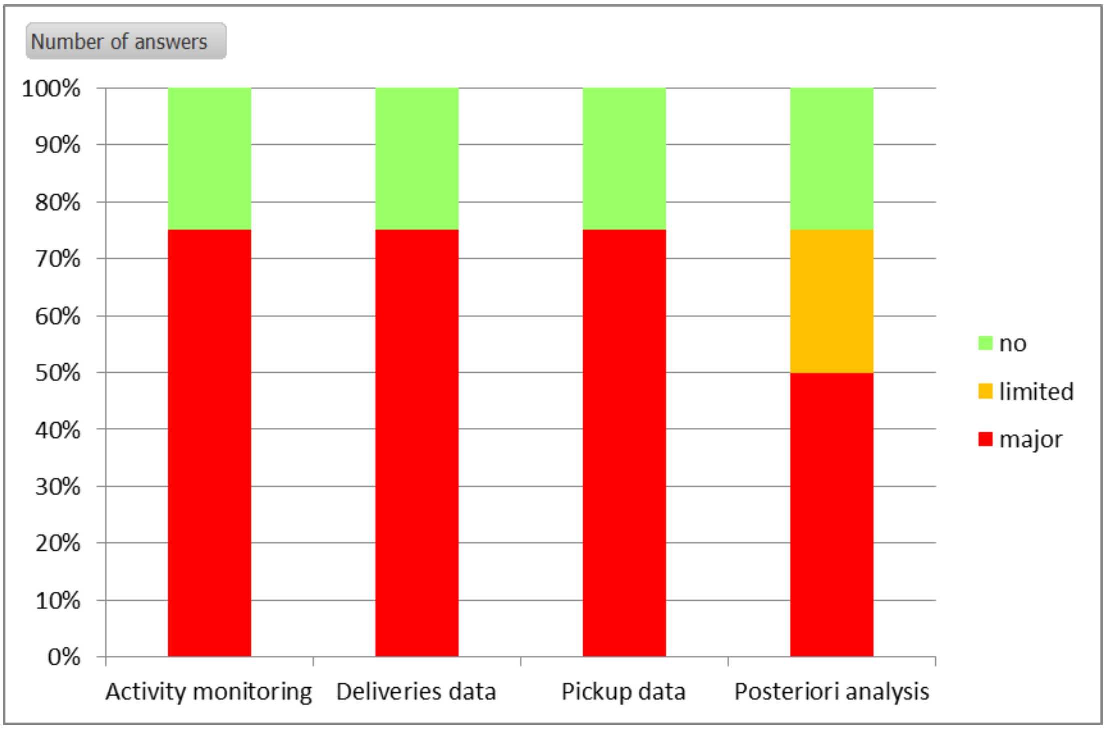


Figure 21: Pilots assessment of datasets confidentiality sensitiveness

5.4 Pilots position towards open-data release

Regarding the possible release of their data, the pilots were asked their position in case of release of the data without any modification (raw data release) and their position in case of release of selected and altered data (e.g. anonymised data).

The pilots are globally not in favour of releasing raw data, as shown in Figure 22.

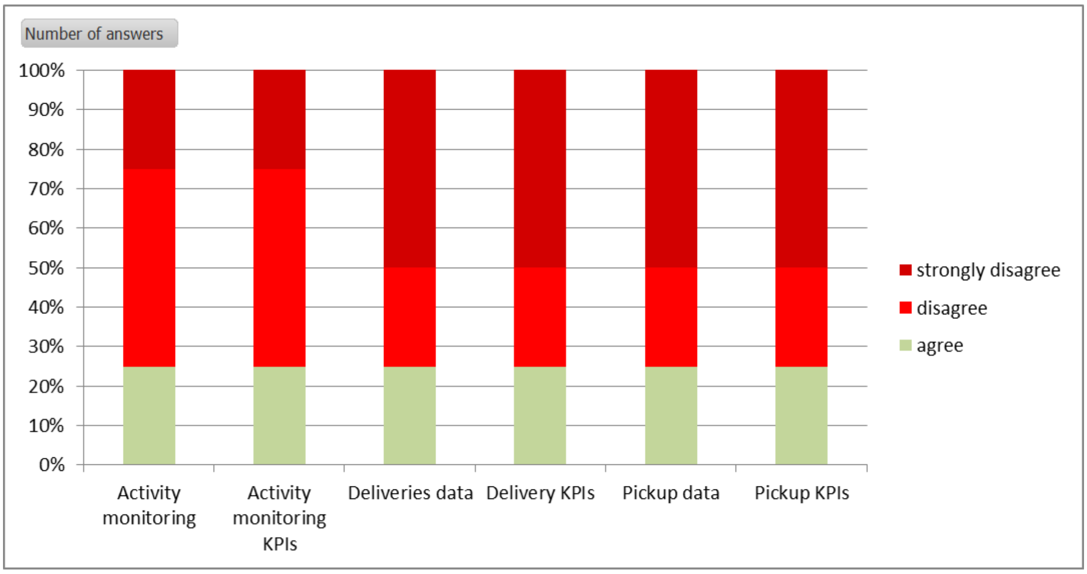


Figure 22: Pilots position towards release of raw data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

This position is more mitigated if the data are filtered (selection of data, alteration or anonymisation of sensitive data) before release (see Figure 23).

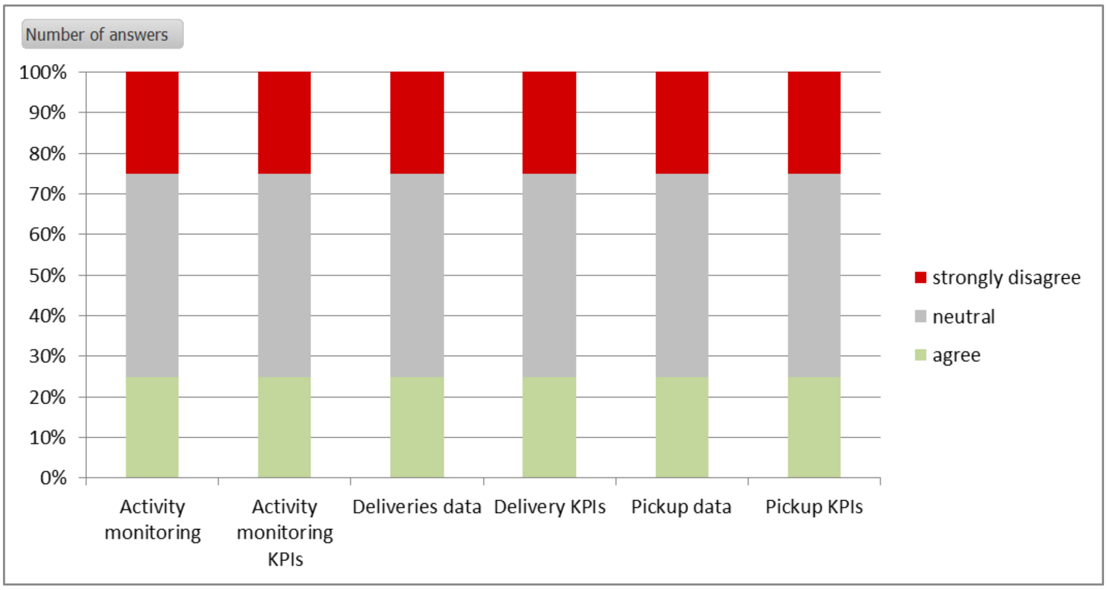


Figure 23: Pilots position towards release of filtered data

5.5 Pilots position on usage conditions for released data

Last the pilots were asked their position on the usage conditions or usage restrictions to set on data release (see Figure 24). Conditions commonly found for the attribution, distribution and usage of open data were proposed.

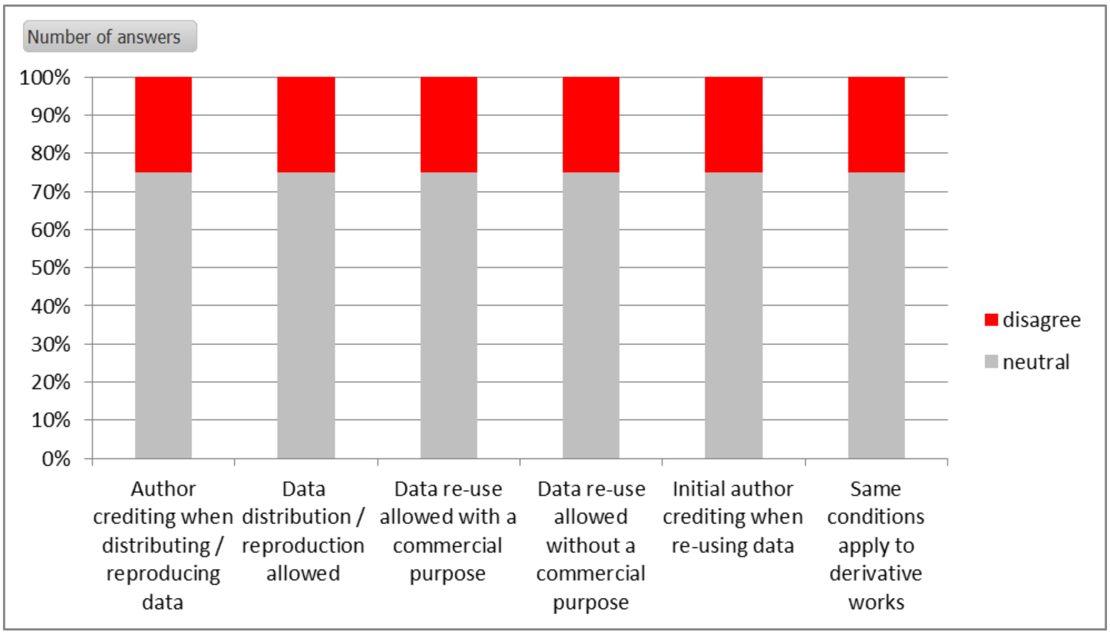


Figure 24: Pilots position on released data usage conditions

The answers are uniform for all conditions proposed and do not show a clear positioning towards a particular option. This means that further explanations of conditions will be required before making any decision on such conditions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

6 Recommendations for releasing data as open data

In the light of the data quality analyses, and the open data survey results, conditions to open the data collected in the work package have been established. Note that the following sections are only recommendations. The final decision to open data shall be taken by the project management board with the formal agreement of the data owners.

6.1 Possible actions to mitigate / reduce risks

Based on the current situation and advices expressed by the project stakeholders, it would only be possible to release one out of the 4 pilot's dataset as open-data. The project should and could take several actions to mitigate and or reduce the risks noted by the pilots in order to be able to release more datasets.

The identifying data that is subject either to confidentiality or to risk exposition could be anonymised by applying several techniques (such as encryption). These techniques cannot guarantee a total absence of identification. Indeed, the detailed analysis and crossing of the project datasets with other datasets (publicly available or not) may lead to an indirect identification.

The data quantifying facts observed during the data collection and that are subject either to confidentiality or to risk exposition could be altered with controlled random noise in a way that reduce risks but still allow to perform analyses on the data.

The data whose interest for research is limited but that are still be subject to risk exposition or confidentiality could be simply excluded from the data release.

After this set of actions, risks shall be re-assessed among the pilots to determine the final set of data to release. This second assessment can be the opportunity to clarify the other risks mentioned by the pilots.

6.2 Intellectual property rights

Related to the dataset collected under a specific agreement from a construction company not member of the project consortium, the project consortium should try to find a contradictory agreement with the data owner to allow the inclusion of the data in the release. If such agreement is not possible then the data shall be excluded from the release.

For the remaining datasets the pilots, as owners of the intellectual property rights on their data, should explicitly agree on the data to release and on the license to apply on their data. This can be formalised in a specific agreement among the consortium if not already foreseen in the grant agreement.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

6.3 Usage conditions for the potential dataset to open

Based on the pilots' feedback from the open-data survey and on the objectives of the SUCCESS project as stated in the proposal, we suggest to apply widely known and commonly accepted rights and conditions on the potential release dataset. The creative commons is a good basis for our recommendations[3].

Attribution: we recommend to apply an attribution principle on the released data so that the project and partners involved in the data collection can be credited for their work.

Non-commercial use. As the initial aim of the project towards open-data is to allow an usage for further research, one could consider applying a principle of non-commercial use. This principle would allow an individual or non for profit usage of the data while preventing users of the released data to include such data in their commercial services and products. One can also imagine a dual licensing schema where the public licence is free (no cost) but has a non-commercial usage restriction, while a commercial licence is available at some cost and allow the inclusion of the data in commercial services and products. The cost of the commercial license can be either based on a one-off cost or on a royalty schema. Both amounts (of fixed cost or royalty rate) need to be carefully determined by further analysis.

Non derivative works. It does not really make sense to prevent derivative works based on released open data. Indeed, the purpose of releasing data is to allow data combination and cross analysis with other datasets. Therefore we suggest not to apply a non-derivative clause on the released data.

Share alike. It seem relevant to make derivative works based on open-data as available as were the initial data, at least for public licences. Therefore we recommend to apply a share-alike principle on the public licence.

As a summary our proposal is a dual licence:

a) By default, a **public licence**, **free** of charge, allowing **non-commercial usage** under **share-alike** and **attribution** principles such as the creative commons **CC BY-NC-SA 4.0**[4]
b) On request and formalized in dedicated agreements, a **commercial licence**, at a specific **cost** (to be determined), allowing **commercial usage** under an **attribution** principle, such as the creative commons **CC BY 4.0**[5]

[3] https://creativecommons.org/share-your-work/licensing-types-examples/
[4] https://creativecommons.org/licenses/by-nc-sa/4.0/
[5] https://creativecommons.org/licenses/by/4.0/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

6.4 Location of the potential dataset to open

By default the datasets could be published on the SUCCESS project website as comma separated text files.

In order to make them more visible, those datasets could be registered on national portals for open-data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

7 Conclusions

We have detailed the actions that have been taken in order to ensure the quality of the most important dataset collected by the project during the second work package. We have shown that data are of enough quality for the purpose of KPI computation for the As-Is analysis.

We have analysed further the quality of the date for the purpose of simulation in order to make sound decisions (notably regarding data samples) in the subsequent tasks and work packages of the project. In summary the main focus points for subsequent tasks of the project are:

- The 1-to-many relationship between the supplier and the address of loading /unloading locations which requires clarification on the share of material that comes / goes to each address depending on the supplier.
- The varying size and composition of the lists of most delivered and collected material depending both on the pilot and on the criteria used to determine the "*most*" (transported weight or number of trips).
- In general, vehicles' load is limited by the volume capacity rather than the weight capacity.

We have identified that if collected data are of good quality enough to be published as open data, pilots as owner of the data are not yet ready to release their data due to several risks they have identified. We have proposed a series of measures to decreases these risks as well as a possible set of licenses to apply to the released data.

The project management board and the impacted stakeholders have to decide on these actions and then to coordinate in order to ensure the release of a dataset as open, as complete and as representative as possible. One of the four dataset could be released quite quickly pending that IPR are clarified and usage conditions are defined and agreed. This first release could be used as an example to trigger the decision of the remaining pilots.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Appendix 1 – Implemented data quality controls

control code	criterion	Purpose	data	control
1	completeness	KPI computation (PM, CO2 equivalent)	Material weight	value is input
2	completeness	KPI computation (Distance from the production to the construction site)	Address of the loading location	value is input
3	completeness	KPI computation (Travel time outside the city centre)	Starting time from the loading location	value is input
4	completeness	KPI computation (Travel time outside the city centre , Travel time inside the city centre)	Arrival time in urban area	value is input
5	completeness	KPI computation (Travel time inside the city centre, Truck punctuality, Truck waiting time outside the site)	Arrival time near the construction site	value is input
6	completeness	KPI computation (Truck waiting time outside the site, Truck waiting time inside the site, Congestion on construction site)	Arrival time at construction site	value is input
7	completeness	KPI computation (Truck punctuality, Construction site punctuality)	Planed delivery time 1	value is input

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

control code	criterion	Purpose	data	control
8	completeness	KPI computation (CO2 equivalent, PM, Distance from the production to the construction site)	Come back empty or loaded	value is input
9	completeness	KPI computation (Distance from the production to the construction site)	Adress of final come back	value is input
10	completeness	KPI computation (PM, CO2 equivalent)	Vehicle type	value is input
11	completeness	KPI computation (CO2 equivalent, PM)	Euroclass	value is input
12	completeness	KPI computation (PM, CO2 equivalent)	Fuel type	value is input
13	completeness	KPI computation (CO2 equivalent, PM)	Net weight	value is input
14	completeness	KPI computation (PM, CO2 equivalent)	Gross vehicle weight rating (GVWR)	value is input
15	completeness	KPI computation (Congestion on construction site)	Surface vehicule	value is input
16	completeness	KPI computation (Construction site punctuality, Loading / unloading time, Truck waiting time inside the site)	Starting time of unloading	value is input
17	completeness	KPI computation (Congestion on construction site, Loading / unloading time)	Ending time of unloading	value is input

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

control code	criterion	Purpose	data	control
18	completeness	KPI computation (Distance from the production to the construction site)	Address of the previous delivery 1	value is input and "direct or Round trip" is "Roundtrip"
19	completeness	KPI computation (Distance from the production to the construction site)	Address of next delivery 1	value is input and "direct or Round trip" is "Roundtrip"
20	completeness	KPI computation (Distance from the production to the construction site)	Round trip or Direct trip	value is input
21	completeness	KPI computation (Distance from the production to the construction site)	Address of the previous delivery 2	value is input and "direct or Round trip" is "Roundtrip" and value is input for "address of previous/next delivery1"
22	completeness	KPI computation (Distance from the production to the construction site)	Address of next delivery 2	value is input and "direct or Round trip" is "Roundtrip" and value is input for "address of previous/next delivery1"
23	accuracy	KPI computation (Distance from the production to the construction site)	Address of the loading location	value returns a location when geocoded
24	accuracy	KPI computation (Distance from the production to the construction site)	Address of the previous delivery 1	value returns a location when geocoded
25	accuracy	KPI computation (Distance from the production to the construction site)	Address of the previous delivery 2	value returns a location when geocoded

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

control code	criterion	Purpose	data	control
26	accuracy	KPI computation (Distance from the production to the construction site)	Address of next delivery 1	value returns a location when geocoded
27	accuracy	KPI computation (Distance from the production to the construction site)	Address of next delivery 2	value returns a location when geocoded
28	accuracy	KPI computation (Distance from the production to the construction site)	Adress of final come back	value returns a location when geocoded
29	consistency	KPI computation (CO_2 equivalent, PM)	Material weight	GVWR ≥ Net weight + Material weight
30	consistency	KPI computation (PM, CO_2 equivalent)	Net weight	GVWR ≥ Net weight + Material weight
31	consistency	KPI computation (CO_2 equivalent, PM)	Gross vehicle weight rating (GVWR)	GVWR ≥ Net weight + Material weight
32	consistency	KPI computation (Travel time outside the city centre)	Starting time from the loading location	Starting time from the loading location > Arrival time in urban area
33	consistency	KPI computation (Travel time outside the city centre , Travel time inside the city centre)	Arrival time in urban area	Starting time from the loading location > Arrival time in urban area

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

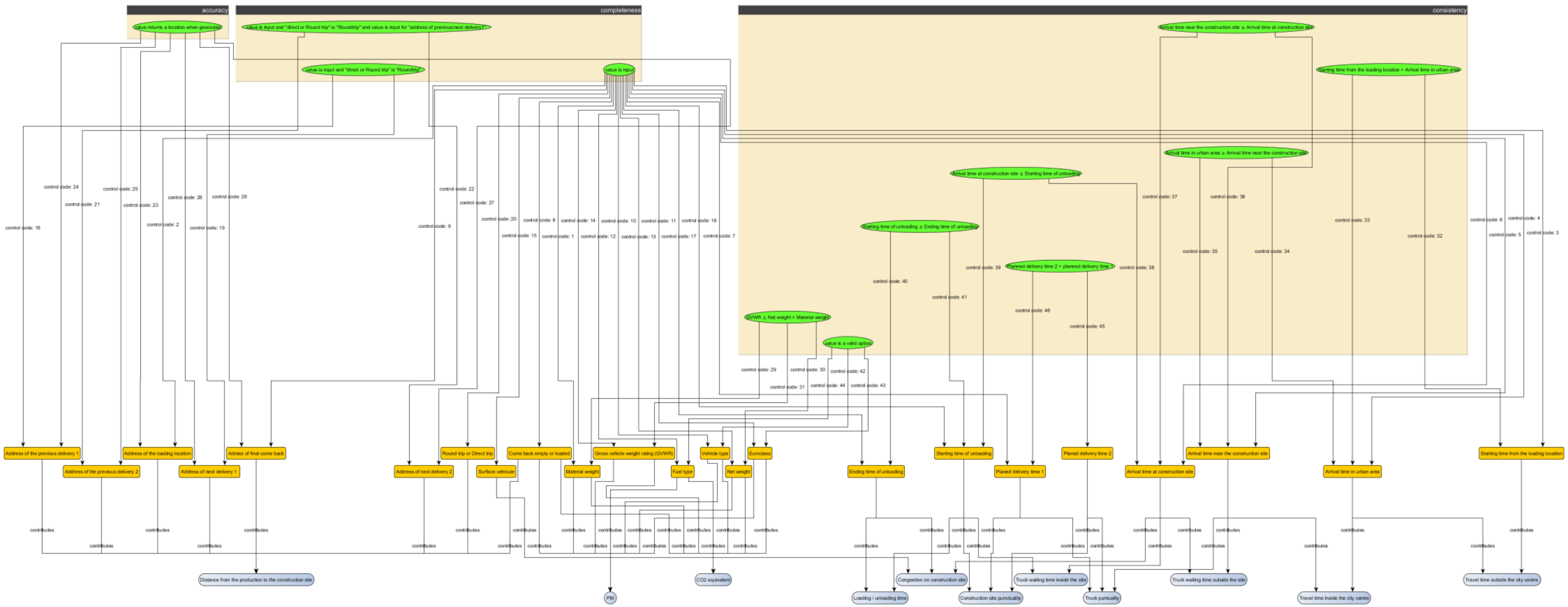
control code	criterion	Purpose	data	control
34	consistency	KPI computation (Travel time inside the city centre, Travel time outside the city centre)	Arrival time in urban area	Arrival time in urban area ≥ Arrival time near the construction site
35	consistency	KPI computation (Truck punctuality, Travel time inside the city centre, Truck waiting time outside the site)	Arrival time near the construction site	Arrival time in urban area ≥ Arrival time near the construction site
36	consistency	KPI computation (Truck punctuality, Travel time inside the city centre, Truck waiting time outside the site)	Arrival time near the construction site	Arrival time near the construction site ≥ Arrival time at construction site
37	consistency	KPI computation (Congestion on construction site, Truck waiting time outside the site, Truck waiting time inside the site)	Arrival time at construction site	Arrival time near the construction site ≥ Arrival time at construction site
38	consistency	KPI computation (Truck waiting time outside the site, Congestion on construction site, Truck waiting time inside the site)	Arrival time at construction site	Arrival time at construction site ≥ Starting time of unloading
39	consistency	KPI computation (Loading / unloading time, Construction site punctuality, Truck waiting time inside the site)	Starting time of unloading	Arrival time at construction site ≥ Starting time of unloading

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

control code	criterion	Purpose	data	control
40	consistency	KPI computation (Loading / unloading time, Congestion on construction site)	Ending time of unloading	Starting time of unloading ≥ Ending time of unloading
41	consistency	KPI computation (Loading / unloading time, Construction site punctuality, Truck waiting time inside the site)	Starting time of unloading	Starting time of unloading ≥ Ending time of unloading
42	consistency	KPI computation (PM, CO2 equivalent)	Vehicle type	value is a valid option
43	consistency	KPI computation (CO2 equivalent, PM)	Euroclass	value is a valid option
44	consistency	KPI computation (PM, CO2 equivalent)	Fuel type	value is a valid option
45	consistency	KPI computation (Truck punctuality, Construction site punctuality)	Planed delivery time 2	Planned delivery time 2 > planned delivery time 1
46	consistency	KPI computation (Truck punctuality, Construction site punctuality)	Planed delivery time 1	Planned delivery time 2 > planned delivery time 1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Appendix 2 – KPI Data Controls Mapping (full view)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Appendix 3 – List of most delivered and collected materials per site

This appendix lists for each site the most delivered and most collected materials in terms of number of deliveries or pickups and in terms of weight delivered. The materials listed constitute 80% of the materials delivered for each criterion: number of deliveries and weight (see the explanation of the Pareto analysis in section *4.1.3 Material*

Completeness

We check the number of times the material name is input per construction site deliveries.

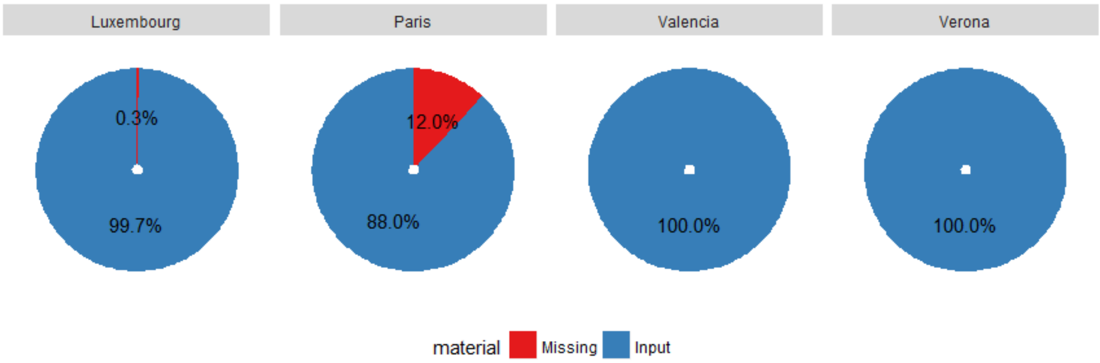


Figure 9: Completeness of material name per pilot site

We notice that **for 3 pilot sites, the material name can be used to identify the material delivered**. **For the last site**, the rate of incompleteness is quite higher and **more than 10% of the deliveries do not specify the type of material delivered**.

Materials most delivered). Note that for Paris, the weight of the delivered material was never input, therefore the weight criterion has not been analysed. The weight is in kg.

7.1 Deliveries

Table 0.1: Most delivered materials on Luxembourg site

Number of deliveries		Rank	Delivered weight	
Top materials	Number		Top materials	Weight
concrete reinforcement steel	39	1	plaster tiles	397121
plaster tiles	25	2	concrete reinforcement steel	220632
scaffolding	20	3	precast walls	156300

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Number of deliveries		Rank	Delivered weight	
Top materials	Number		Top materials	Weight
precast walls	16	4	alveolate precast slabs	153578
facade insulation plate	13	5	cinderblock	136400
windows	13	6	precast balcony	100073
basement insulating	11	7		
floor tile	11	8		
precast balcony	11	9		
alveolate precast slabs	9	10		
cement for cope	9	11		
cinderblock	9	12		
curtain wall	9	13		
cinderblock	8	14		
precast stairs	7	15		
formwork material	6	16		
foamed concrete	5	17		
plaster for coating	5	18		
sand	5	19		
boilers	4	20		
cable path	4	21		
door	4	22		
isocorb	4	23		
prefabricated rebar connection	4	24		
electrical component	3	25		
kitchen	3	26		
metal cladding	3	27		
mortar	3	28		
precast slab	3	29		
electrical cable	2	30		
fire doors	2	31		
lift	2	32		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Table 0.2: Most delivered materials on Paris site

Number of deliveries		Rank
Top delivery materials	**Number**	
pipes	155	1
plasterboard	118	2
concrete bloc	89	3
coating	87	4
sand	48	5
sheath	44	6
cable	41	7
partition wall	39	8
insulation	38	9
window pane	36	10
door	32	11
scaffolding	26	12
treste	26	13
electric material	25	14
fiberglass	23	15
door frame	20	16
insulating	19	17
paint	19	18
bench grinder	18	19
gravel	18	20
light fitting	18	21
rail	18	22
tiling	18	23
cement	17	24
sheet metal	17	25
staff	16	26
coat	15	27
scrap	15	28
carpet	14	29
blind	12	30
false floor	11	31

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Number of deliveries		Rank
Top delivery materials	**Number**	
chassis	10	32
slab	10	33
gaz bottle	9	34
panel	9	35
parapet	9	36
ventilator	9	37
consumable	8	38
gable	8	39
material	8	40
palette	8	41
plaster	8	42

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Table 0.3: Most delivered materials on Valencia site

Number of deliveries		Rank	Delivered weight	
Top materials	Number		Top materials	weight
concrete	572	1	brick	448027
brick	31	2	aggregates	230560
		3	gravel	110700
		4	tile	109806
		5	concrete	100844
		6	concrete curbs	76508
		7	panels	73555.14
		8	dry mortar	73130
		9	big bag sand	71555
		10	corrugated tube	57369.33
		11	solid brick	26406
		12	concrete. sand. plaster and porcelain floor	24400
		13	asymmetric cone. ring and wood palette	23790
		14	dry concrete	23520
		15	ondufilm terracotta. rastrel pvc. red washers. spiral nail 8cm.	23277.84
		16	concrete and big bag sand	23060

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Table 0.4: Most delivered materials on Verona site

Number of deliveries		Rank	Delivered weight	
Top materials	Number		Top weight materials	weight
concrete	22	1	concrete	477600
plasterboard panels	17	2	plasterboard panels	468000
insulated glazing	12	3	insulated glazing	115630
elevators	11	4	stoneware tiles	104400
stoneware tiles	10	5	metallic cantilever roof	84000
air handling unit	9	6	resilient floor covering	61000
fire doors	9	7		
plasterboard structure	9	8		
fuse boxes	7	9		
resilient floor covering	7	10		
glue for resilient floor covering	6	11		
plasterboard panels and structure	6	12		
plasterboard panels for false ceiling	6	13		
metallic cantilever roof	5	14		

7.2 Pickups

Table 0.5: Most collected materials on Luxembourg site

Number of pickups		Rank	Collected weight	
Top materials	Number		Top materials	Weight
industrial waste	36	1	industrial waste	27560
treated wood waste	21	2	treated wood waste	26300
residual waste	19	3	plaster waste	22660
formwork	8	4	rond à béton	20980
plaster waste	8	5		

Table 0.6: Most collected materials on Paris site

Number of pickups		Rank	Collected weight	
Top materials	Number		Top materials	Weight

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

waste	320	1	waste	806340
wood	189	2	mixed rubble	590100
mixed rubble	119	3		

Table 0.7: Most collected material on Valencia site

Number of pickups		Rank	Collected weight	
Top materials	Number		Top materials	Weight
inert and non-hazardous waste	376	1	inert and non-hazardous waste	9024000

Table 0.8: Most collected material on Verona site

Number of pickups		Rank	Collected weight	
Top materials	Number		Top materials	Weight
wood	24	1	mixed packaging	97100
mixed packaging	18	2	cement	97000

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.

Table of figures

Table of tables

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 633338.