CAMERA

D4.1:
Gaps, bottlenecks and results: the methodology

H2020 CAMERA – COORDINATION AND SUPPORT ACTION FOR MOBILITY IN EUROPE: RESEARCH AND ASSESSMENT

CAMERA Consortium:

1	innaxis	INNAXIS (INX)	Contact: Samuel Cristobal sc@innaxis.org
2	UNIVERSITY OF WESTMINSTER	University of Westminster (UoW)	Contact: Andrew Cook cookaj@westminster.ac.uk
3	EUROCONTROL	EUROCONTROL (ECTL)	Contact: Peter Hullah peter.hullah@eurocontrol.int
4	Bauhaus Luftfahrt Neue Wege.	Bauhaus Luftfhart s. V. (BHL)	Contact: Annika Paul annika.paul@bauhaus-luftfhart.net
5	deepblue CONSULTING & RESEARCH	Deep Blue srl (DBL)	Contact: Micol Biscotto micol.biscotto@dblue.it

Document number: D4.1

Document title: : Gaps, bottlenecks and results: the methodology

Version: 1.0

Status: Final

Work Package: 4

Deliverable Type: PU

Delivery date: April 2019

Responsible Unit: University of Westminister

Contributors: Gerald Gurtner (UoW), Andrew Cook (UoW), Damir Valput (INX)

This project has received funding from the European Union's Horizon 2020 research And innovation programme under grant Agreement n°769606

Table of Contents

1 Introduction

CAMERA WP4 analyses the output of the mobility model (WP3), within the performance framework defined in WP2, in order to identify gaps and bottlenecks in European mobility research that may impede the high-level goals set, for example, by Flightpath 2050.

In particular, this deliverable presents the methodology that CAMERA will follow to produce an overview of the European transportation research field. This methodology is based on the analysis carried out, and, planned in WP3, including the output of the cluster analysis presented in D3.1 and D5.3.

The aim of the methodology herein is to produce a high-level view of the mobility field by highlighting its main strengths and weaknesses. To do this, a cross-analysis is performed using topics from the cluster analysis and the layers defined by CAMERA in D2.1.

Experts will prioritise the sub-topics, using high-level objectives. Further analyses will be performed on selected, key projects, using manual and automatic mechanisms.

2 Meta-analysis

2.1 Layer and topic alignment

The macro-modelling of European research projects, as described in D3.1, results in two independent project classifications. The semi-supervised modelling methodology based on Latent Dirichlet Allocation (Blei et al., 2003) automatically extracts topics from the set of projects and produces a number of word clouds that describe these generated topics. This results in a topic distribution against the automatically extracted topics for each project from the data set. The pie chart in Figure 1 illustrates a potential topic distribution and assignment of projects to its most relevant topics (the pie chart is obtained according to the preliminary results: see Mobility Report 1 for more details).

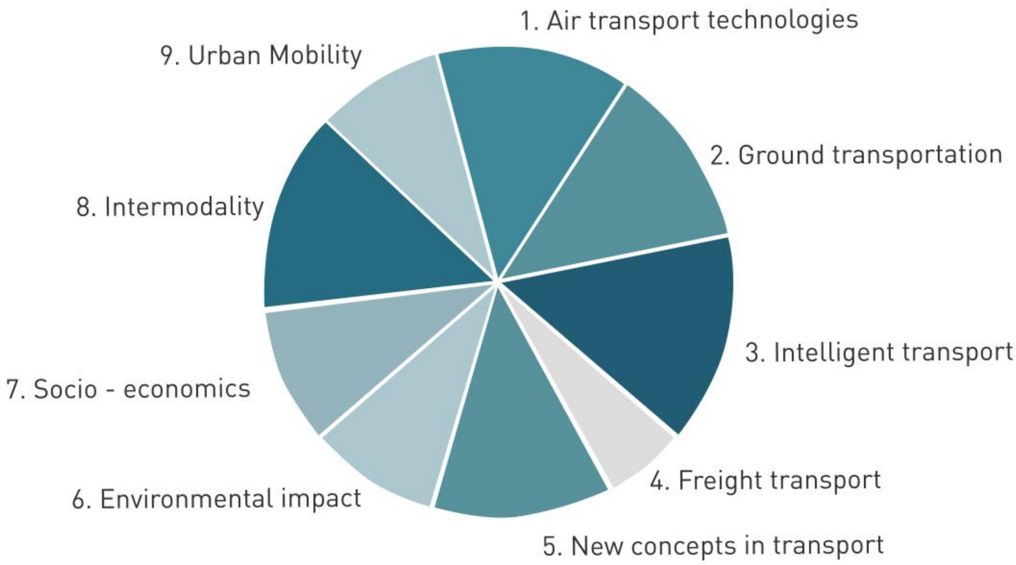

Fig.1. Topic distribution and assignment of projects

Furthermore, projects are also assessed against the mobility layers previously defined in the CAMERA performance framework (CAMERA D2.1, 2018), whereby each layer expresses a high-level objective of European mobility research. This assessment also relies on semantic similarity metrics and text-mining methods, providing a probabilistic classification of the projects against the five layers. In other words, using semantic similarity metrics, each project is assessed against the objectives of each layer, using textual summaries of projects and descriptions of layer objectives. For illustration purposes, a heatmap such as that shown in Figure 2 can be used to visualise the similarity scores between projects and layers, translating the semantic similarity scores into Z-scores (Mendenhall, 2007).

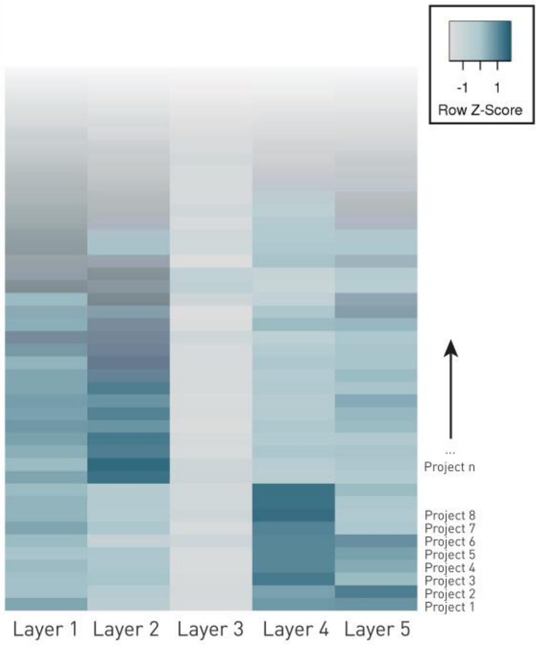

Fig.2. Similarity scores between projects and layers

In order to establish the relationship between the topics and the layers, the results of the two modelling methodologies above can be used. Using the probabilistic distributions generated, each independent of the other, a score can be assigned to each topic-layer pair that quantifies their correlation.

The score can be derived using a cumulative probability obtained from the two probabilistic distributions (Blei, 2012). To have a single score for each topic-layer pair, the cumulative probabilities need to be assessed and summed across all the mobility projects.

This results in a matrix of scores that aligns topics with layers. Other variations of score definition can be considered if needed, such as giving a higher weight to the dominant topic for each layer, using topic ranking instead of exact probability distributions, etc.

In addition to the probabilistic distribution, the scoring described above could be further informed by the results of an expert-based assessment, as described in Section 3. Such addition of the expert-based assessment to the algorithm-based score can be seen as a regularisation method, and thus serve as a way to reduce the potential biases of the algorithmic assessment.

2.2 Gaps and bottlenecks

Other analyses can be performed on the output of the cluster analysis, with or without the additional knowledge related to the layers.

2.2.1 Topic-layer analyses

The distributions of projects per topic is a good indicator of the balance between the topics. Over-representation or under-representation can be detected using statistical methods, for instance using a Fisher test or a permutation test. These tests check if, under a null hypotheses of a random distribution of topics per layer, one of the layers has a significant deviation from the distribution, imputable to something other that randomness.

Under-represented topics are likely to represent under-funded areas (i.e. topics having fewer projects associated to them). However, it is even more important to detect under-funded areas in terms of layers, which follow (non-rigidly) more high-level objectives for European mobility.

The analysis explained in Section 2.1 between topics and layers allows CAMERA to have an indirect link between projects and layers. As a result, one can derive over- and under-representation for layers too, in the same way as described for the topics. This analysis gives important information in terms of areas which should be funded, since all layers defined by CAMERA are important for European mobility. Moreover, combining this analysis with the topic analysis could help to better focus the needs for investment.

The comparison between topics and layers *per se* is informative. Indeed, topics can roughly be thought of as 'research that happens in reality' and layers as 'research that should happen'. The difference between them may help with redefining high-level objectives or/and better reorienting current research. Coverage of layers should be fairly uniform, unless there is a good reason to think that one of them needs more or less intensive research. As a result, CAMERA will use different metrics to capture this effect, including variance, and will highlight the outliers. Statistical tests can also be used here, such as permutation or bootstrap tests.

The analysis described in Section 2.1 gives a score for each topic-layer pair. It is expected for each layer to have a high score on several different topics. The relative importance of topics for each layer can be assessed using the intersection-over-union measure, i.e. the Jaccard similarity measure (Jaccard, 1912). Using this measure normalises the obtained scores and gives the probability of each topic being relevant to a layer. Moreover, the 'ownership' of a topic can be considered as shared between different layers. A further analysis is to measure the degree of 'purity' of the topic in each layer, i.e. how much it is *not* shared among layers. Several metrics could be used for this, including the Gini impurity index or information gain. CAMERA will test different metrics and compare the results.

One can also use other metrics to compare the two clusterings (topic- and layer-based), for instance the Rand index (Rand, 1971) or mutual information (Cover, 1991). These metrics allow us to have more detailed information on how consistent the two clusterings are. For instance, if two project are in the same topic, are they also in the same layer?

2.2.2 Spatio-temporal analyses

An important analysis to perform to identify gaps is spatial and temporal analysis. Temporally, it is is important to see whether some subjects (materialised by topics or layers) are trending, and if some subjects are likely to increase in the future. The duration of each project is simple to find out, and one can use a simple graph such as that in Figure 3 to capture these trends.

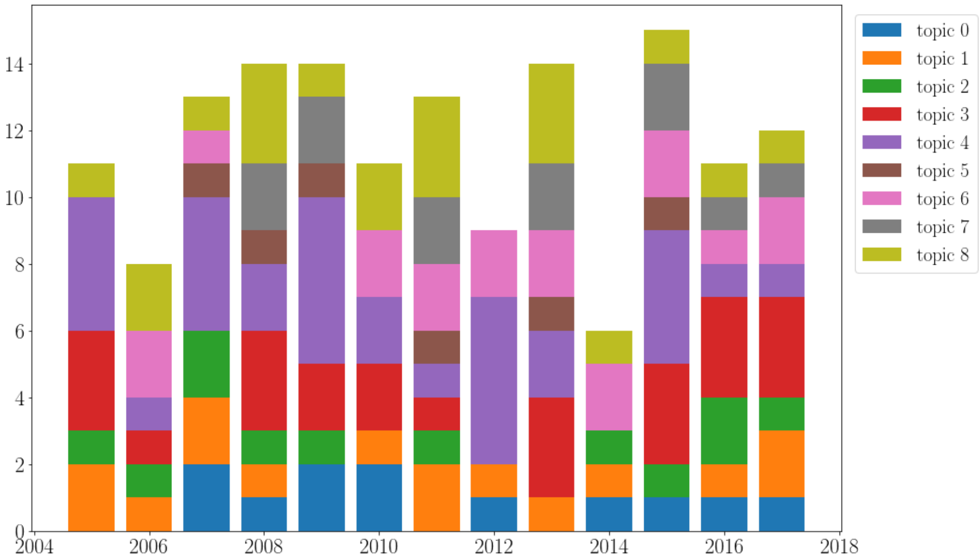

Fig.3. Example representation of the temporal distribution of projects in each topic

More interestingly, the projects may be non-uniformly distributed among countries. Indeed, some countries might tend to specialise in some areas, while others are more generalist. Statistical tests can be used to detect these regularities (such as the Fisher test). Several null hypotheses can be used to test differences, to properly take into account country population, for instance. Under-represented countries in some areas might indicate some lack of funds, or on the contrary, a good specialisation of the research team in other domains, as Figure 4 illustrates (using an example plot only at this stage).

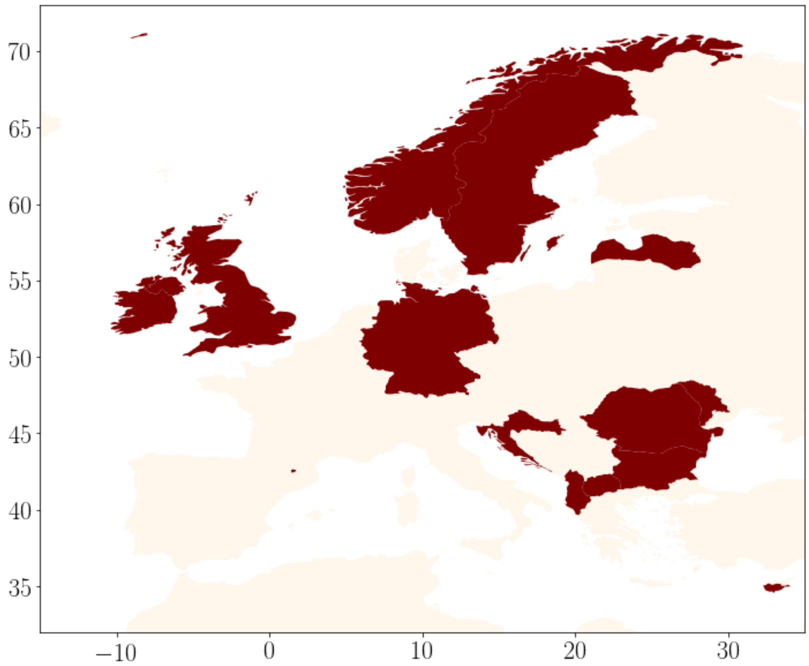

Fig.4. Example representation of countries where an over-representation of projects in a given topic is detected

3 Focused analysis

After the automatic meta-analysis described in Section 2, CAMERA plans to perform some more focused analysis, using first expert judgement and then a quantitative model. CAMERA will deploy expertise within the consortium, the project's Advisory Board, and additional eperts. The Advisory Board is comprised of 15 members from a wide range of stakeholders (e.g. airports, ANSPs, rail industry), thus adding value regarding the feedback that they will provide. Other experts will also later assess the earlier evaluation steps in CAMERA, in particular during a workshop. The output of this assessment will lead to policy recommendations from the CAMERA team, as well as the selection of a small number of projects which will be further analysed by CAMERA's quantitative tool, 'Mercury'. These steps are described below.

3.1 Expert analysis

CAMERA plans to have interactions with experts over three different phases, centered around a workshop. These external experts are composed of two groups, *viz*. the:

- CAMERA Advisory Board;
- (other) attendees at a workshop.

The first group will be asked to perform some tasks before and after the workshop. The workshop will serve as a central event to draw on more knowledge and feedback.

3.1.1 Phase 1 - evaluate objectives and topics

The aim of the first phase is to prepare a prioritisation of high-level objectives for mobility in Europe and assess how they relate to the list of topics generated by the automatic clustering process (and possibly modified by the process described in Section 2). This phase takes place a few weeks before the workshop of Phase 2.

Firstly, the CAMERA team will decide on a list of high-level objectives drawn for instance from the ACARE SRIA. These objectives should be wide enough to encompass mobility concepts that are not directly linked to the air transportation system. The exact number of objectives will be decided upon during this review, but should not be very high, to focus these analyses relatively tightly. Some of these objectives, at least, should be measurable in Mercury for further analysis (see Section 3.2). Even though some of them can be qualitative, most of them should be directly related to the KPIs identified in CAMERA (in D2.1). Some of them can also be quantitative and be suggested as new KPIs.

This list of objectives will then be shared with the Advisory Board. They will be asked to rank them, for example by allocating a fixed amount of (investment/spending) 'credits' between them. They will also be given the opportunity to discard an objective and replace it with one of their choice (either from a published source or based on their opinion, if any omission is perceived).

With this ranking, the Advisory Board will be ready to assess the topics from the automatic clustering that were communicated to them. For each of their top objectives (e.g. the first three), they will be asked to rank the topics that are more likely to deliver significant improvements and work towards this objective.

Moreover, they will be asked whether some topics should be merged and, more generally, if the topics are sufficiently specific but cover enough ground overall. Based on this feedback, the CAMERA team will be able to adjust the topics, for instance merging topics that are too similar.

Brief feedback will be sought regarding the layers themselves. Whilst the layers are fixed in CAMERA, the team is interested in the Advisory Board's view regarding the relative importance of the layers, and degree of interdependency. An evaluation similar to the topic-related task will thus be put to them. This will allow us to adjust some of the null hypotheses used during the meta-analysis of Section 2.

3.1.2 Phase 2 - gaps and bottlenecks (with workshop)

The aim of the second phase of interaction is to examine the overlap between the data-driven topics and the pre-defined layers of CAMERA and to identify areas of interest – where (additional) research investment could be made, for instance. This second phase takes place primarily during a workshop.

The workshop will coordinate the Advisory Board and other external experts (collectively referred to herein simply as the 'experts'). Close cooperation and coordination with ACARE WG1[1] is expected. CAMERA regularly attends this group's meetings, updating it on the project's progress such that the representatives are already familiar with its methodology. ACARE WG1 members thus provide valuable feedback to the project, which represents very useful support. In addition, the expertise of those attending these meetings is highly relevant for CAMERA's scope.

Accordingly, the ACARE WG1 members will be considered as ideal attendees for the workshop. This will last one day and will be organised in interactive sessions where participants can provide specific feedback. If possible, this will be organised together with an ACARE WG1 meeting to facilitate the attendance of the members at both events.

After a brief review of phase 1, the CAMERA team will present a consolidated list of the high-level objectives resulting from this phase. Participants will then be divided into groups. Groups could be homogeneous in terms of expertise (e.g. one type of stakeholder, such as 'airlines') or (partially) mixed, to foster discussion. The exact number of groups will be decided just prior to the workshop based on the actual attendance anticipated, but will typically include fewer than eight persons each (to maintain focus within each group). The main objectives of the groups will be to prioritise some areas of funding, which are of importance to reach the objectives, based on the topics and layers. The exact methodology of the workshop will be fixed at a later stage and is subject to revision, but is outlined below.

In the first step, the participants will be presented with a *simplified* summary of the outputs of Section 2.1 (as per Figure 5), for example identiying the strongest topic/layer overlaps (in blue), and asking the experts to complement those with their own expectations (in selected areas where they have sufficient knowledge).

Clearly, the experts cannot be expected to *quantify* these 45 overlaps, such that this task would be used more as a qualitative and complementary step to that of Section 2.1, potentially identfying any algorithmic biases or errors in that process (i.e. a regularisation method), and to help focus the experts' minds on the intersection/overlaps, in readiness for the second step.

In this second step, participants will be asked to use objectives stickers on a white board, and to distribute credits according to the *desirability of investing* in the various intersections, for example of topic 1 and layer 1 (Figure 6; top-left hand cell, shown in green – implying higher desirability for investment).

[1] Advisory Council for Aviation Research and Innovation in Europe, Working Group 1 on Meeting Societal and Market Needs.

Topic/ layer	1	2	3	4	5	6	7	8	9
1									
2									
3									
4									
5									

Fig.5. Further qualifying overlaps between topics and layers

Topic/ layer	1	2	3	4	5	6	7	8	9
1									
2									
3									
4									
5									

Fig.6. Quantifying investments at intersections between topics and layers

The division into groups (potentially representing distinct stakeholder interests) allows cross-examination, which leads to the discovery of biases and allows us to address the possible lack of consensus on given issues. A plenary session will then be organised to synthesise the results and form some recommendations for future research initiatives.

The synthesis may be made via different methods, including decision by mutual consent, vote to majority, average scoring etc, which will be decided at a later stage. The synthesised results in the second table will also be used to modify the procedure described in Section 2.1, to refine the results.

3.1.3 Phase 3 - further analysis

The last phase begins with the CAMERA team selecting a few projects in the most promising areas detected in the previous phase. The selection will be made based on the following criteria:

1(a): identify best-in-class projects in given cells;

or

1(b): how representative (typical) the project is, of other projects in the same cell;

and

2. how easy it is to further analyse the project with Mercury (see Section 3.2.2).

The first criterion will be assessed judgmentally by the CAMERA team, on the basis of the rigour of the approach described (in the project reporting) and qualities *such as* the level of quantification, degree of evidence-led reporting, depth of state-of-the-art reviewing, etc., deployed in the project. Whilst this will necessarily be fairly high-level, it will be used to identify best-in-class approaches and example projects to consider further, rather than to report upon or validate the projects.

The representativeness of the project (1(b)) will be assessed using the KPAs addressed in the project, as well as its scores derived from the cluster analysis. Depending on the feasibility of obtaining sufficient evidence to formulate a view on 1(a) and 1(b), for example the number of projects available in a given cell and the accessibility of further reporting, *either* criterion 1(a) *or* 1(b) will be used. The second criterion (2) will be assessed using the division into classes, as explained in Section 3.2.2.

The second step in this phase will be for CAMERA to assess the project's impact on the European transportation system. This will be done by extracting their contribution in terms of the KPIs defined with CAMERA in D2.1.

Moreover, a quantitative analysis will be run on the concepts proposed by the projects using an integrated model (Mercury), which will be able to assess the impact of the concepts overall (possibly as a function of uptake, or based on assumed traffic conditions, etc.) on the European transportation system. The output of the model will be analysed as described in Section 3.2.2.

CAMERA will summarise these findings and share them with the experts from phase 2. The final stage of the interaction with them will comprise analysing the differences between the Mercury output and projects and judging which factors give rise to such differences. This is important to pave the way for a better consolidated validation/evaluation framework regarding the KPIs to be used, and also the hypotheses and/or baselines used by the projects.

3.2 Quantitative Analysis

3.2.1 Quantitative model – Mercury

The tool CAMERA will use is Mercury. It has been developed by the University of Westminster and Innaxis over several years. Several versions have been deployed in SESAR projects, including POEM, ComplexityCosts, Vista, and Domino.

The simulator features a highly detailed model for passenger mobility, tracking individuals from door to door, taking into account particular connections at airports, rebooking during disruption, and airport access and egress times.

The latest version of the model has been designed as a modular, agent-based simulator where it is reasonably easy to add and remove agents, or modify their behaviour. The model heavily depends on data, in particular to create realistic delay distributions for flights and passengers.

Detailed features of Mercury include:

- data-driven mesoscopic approach, stochastic modelling;
- individual passenger, door-to-door itineraries;
- individual flight plan generation;
- Regulation 261/2004 cost factors included (passenger care and compensation requirements);
- disruption and cancellation modelling;
- stochastic wind generation;
- airline decision-making rules and mechanisms for delay recovery and delay cost management, including wait-for-passenger policies;
- flight prioritisation strategies (closely related to SESAR Solutions, SESAR (2019)).

3.2.2 Analyses planned with Mercury

In deliverable D2.1, CAMERA has divided projects into three classes:

- **Class 1**: can be fully quantified by Mercury, provided the data requirements are met;
- **Class 2**: could be partially estimated by Mercury (e.g. through a relationship between a Class 1 and Class 2 KPI in the literature) or could be easily implemented by extending an existing functionality;
- **Class 3**: cannot currently be estimated by Mercury, e.g. difficult to implement due to data restrictions and/or calculation/estimation complexity.

During phase 3 of the interaction with experts, the CAMERA team will select 2 or 3 projects belonging to the most interesting areas flagged during phase 2. These projects will ideally be in class 1, or in class 2 if no class 1 projects are available in these areas. The team will then implement some of the changes described in the project documents, in Mercury. These changes will be interchangeable, one by one, or conjointly.

The results collected from the Mercury output are microscopic itineraries followed by each individual passenger. Different types of analyses are possible on the output. CAMERA will be mainly focused on measuring aggregated KPIs. These KPI values will depend on the changes introduced, as informed by the projects. As a result, CAMERA will be able to identify:

- interactions between concepts from different projects, i.e. the independent results c.f. the jointly-modelled ones;
- trade-offs between KPIs: in particular, some concepts might be designed to enhance one KPI but also have an adverse effect on others (indeed, quite possibly not measured by the original project).

Where targets are available, gap analyses will also be performed. In particular, major differences between the target and impact of the project in some KPIs will be highlighted. This will be done to emphasise some investments required and their magnitude (where possible).

4 Conclusions and next steps

This deliverable has described the main steps that CAMERA will undertake to analyse the outputs of WP3 (e.g. topic modelling), in the context of WP2 (e.g. the CAMERA performance framework).

After a round of automatic tests and analyses, experts will be at the centre of various assessments, which will lead to concrete recommendations.

The next steps will involve the actual analysis of the output of WP3 and the outcomes of the expert consultations, which will be primarily reported in:

- Deliverable 3.2 (Mobility research assessment and modelling: implementation);
- Deliverable 4.2 (Qualitative and quantitative assessment implementation); and,
- Mobility Report 2.

5 References

CAMERA D2.1 (2018). *D2.1 Establishment of Performance Framework*. Version: 1.1. CAMERA: Coordination and support Action for Mobility in Europe: Research and Assessment, H2020 project.

Cover, T.M., Thomas, J.A. (1991). *Elements of Information Theory* (Wiley ed.)

David M. Blei (2012). *Probabilistic topic models*. Commun. ACM 55, 4 (April 2012), 77-84. DOI: https://doi.org/10.1145/2133806.2133826

David M. Blei, Andrew Y. Ng, and Michael I. Jordan (2003). *Latent dirichlet allocation*. J. Mach. Learn. Res. 3 (March 2003), 993-1022.

Fisher, Ronald (1954). *Statistical Methods for Research Workers.*

Jaccard, Paul (1912). *The Distribution of the flora in the alpine zone*, New Phytologist, 11: 37–50, doi:10.1111/j.1469-8137.1912.tb05611.

Mendenhall, William; Sincich, Terry (2007), *Statistics for Engineering and the Sciences* (Fifth ed.), Pearson / Prentice Hall, ISBN 978-0131877061

SESAR (2019). SESAR Solutions Catalogue (3rd ed.)

Rand, W (1971), *Objective criteria for the evaluation of clustering methods*. *J Am Stat Assoc* 66: 846–850.