‘Mind the Gap’: Evaluation Tool for the Implementation of Personalization in Passenger Information Systems

Michelle T. van Ardenne[1(✉)], Matej Cebecauer[1], Oded Cats[1,2], and Zhenliang Ma[1]

[1] KTH Royal Institute of Technology, 100 44 Stockholm, Sweden
mtva@kth.se
[2] Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands

Abstract. Providing relevant information is crucial in public transport systems. With the rise of digital passenger information systems (PIS), personalization has emerged as a means to meet passengers’ information needs better. To better understand how personalization can be implemented in PIS, five levels of personalization have been identified in the literature, highlighting varying degrees of system autonomy and passenger involvement. While these levels have advanced the understanding of personalization, their practical application remains limited. This paper builds upon an existing framework of personalization levels. It introduces an evaluation tool composed of distinct performance measures to help PIS developers assess their system’s current personalization level and identify opportunities for technological advancement. The tool enables a comparison of the personalization functionalities against the best practices defined by the personalization levels. The paper further outlines the creation of the tool through functional benchmarking, introduces system behaviors across levels, and evaluates commercial PIS through case studies, offering actionable insights for advancing PIS personalization.

Keywords: Personalization · Passenger information system · Public transport

1 Introduction

Providing relevant information to passengers is essential for the functioning of the public transport system. Passenger information helps passengers understand how the services are operating and plan their public transport trips. With the digitalization of passenger information systems (PIS), passengers currently have access to various technologies that support their information needs when using public transport, including personalization. According to Vredenborg et al. [1], personalization in PIS refers to “the process of adapting a passenger information system to (a group of) travelers based on personal and contextual factors. This adaptation can be implicit and explicit and aims to achieve a particular goal”. In

C. Stephanidis et al. (Eds.): HCII 2025, CCIS 2771, pp. 52–62, 2026.
https://doi.org/10.1007/978-3-032-12764-8_6

practice, the functionalities of personalization can increase the relevance of passenger information for the passenger, as it adapts information to the information needs of the passenger.

However, there is not one approach to personalize passenger information. Van Ardenne et al. [2] identified 5 levels of personalization which showcase different personalization approaches based on various degrees of passenger participation. By structuring personalization in the five levels, various degrees of system autonomy for personalization have been defined.

The creation of the levels of personalization has been the first step in achieving a better understanding of personalization in the context of PIS. To enhance the practical applicability of the levels, a concrete evaluation tool is needed. PIS Developers would benefit from practical guidance on identifying the current personalization level of their PISs and support for advancing to higher personalization levels. Therefore, this paper aims to extend the framework of personalization levels by van Ardenne et al. [2] with distinct performance measures and structure them into an evaluation tool applicable for developers to determine the next operational and technical developments for their PISs.

The rest of the paper is structured as follows: Sect. 2 provides a recap of the personalization levels framework as well as a discussion on benchmarking in information science. Section 3 details how the evaluation tool is created with the use of functional benchmarking. Section 4 presents changes in system behavior across the levels of personalization. Section 5 presents the evaluation of case studies using commercial PIS. Finally, Sect. 6 provides a recap of the paper's findings and outlines directions for future research.

2 Background and Related Work

Five levels of personalization have been defined to showcase different personalization approaches in PIS [2]. The key distinctions between the levels of personalization are the various degrees of passenger participation required to perform the personalization. According to the framework, the degree of passenger participation is determined by the balance between the system's autonomy and the passenger's agency in personalizing the information for the passenger. At lower levels of personalization, the passenger's agency takes precedence, as passengers actively personalize their experience by manually inputting preferences or interpreting the relevance of provided information. Conversely, at a higher level of personalization, the system autonomy becomes the driving force, with advanced PIS leveraging automation to personalize information.

Looking into the levels, the following dynamics can be identified between the passenger's agency and the system's autonomy [2].

– No personalization represents "a baseline state, where the system provides uniform information to all passengers without any form of adaptation." The system's autonomy at this level is non-existent due to the lack of a mobile

application, leaving the passenger to find, filter, and interpret information manually at dedicated places, such as stations or websites.
- Customization refers to "allowing passengers to provide direct and explicit inputs to the system, and can be characterized by the passenger actively contributing to achieving the personalization." The system, on the other hand, shows few signs of autonomy as it remains passive and executes exactly what is requested based on the passenger's input.
- Partial personalization refers to "systems that personalize based on assumptions derived from the general trips because of simplified functionalities or a lack of data on unique travel situations." The system becomes capable of inferring general preferences and characteristics about the passenger from explicit input data provided by the passenger, applying only to the default travel situation of the passenger.
- Conditional personalization refers to "systems that automate the functionalities of personalization for specific, previously recorded travel situations." The system's autonomy increases as the system no longer relies on preferences explicitly stated by the passenger, but on data of historical travel behavior data, containing detailed records of the passenger's travel situations.
- Full personalization refers to "systems that automate all personalization functionalities in all known travel situations and uses external data sources to automate all functionalities for travel situations where travel data is missing, as far as the external data sources reach". The system proactively personalizes information, leveraging its ability to predict and infer passenger preferences from external data sources, such as personal calendars or social media data.

The balance between the passenger's agency and system autonomy is determined through the characteristics of the core functionalities of personalization [2]. Three core functionalities have been identified: (1) the ability to identify the target of personalization, (2) the ability to be aware of travel situations, and (3) the ability to determine the trigger for information provision. The way these core functionalities are designed in the system determines the roles of the passenger and the system in the personalization of information. It is therefore important to identify and understand the system's behavior that ensures that the functionalities are delivered as intended.

Benchmarking is used to identify important system behaviors and compare the performance of these behaviors across one or multiple systems [3]. Benchmarking aims to find the best practices for the different functionalities of the system. It involves several steps to identify system behaviors and compare their performance. Following these steps, benchmarking enables developers to identify areas for improvement and helps to prioritize the resources to implement the improvements by focusing on system behaviors that are most likely to have a significant impact.

3 Benchmarking Approach for the Evaluation of Personalization

To evaluate PIS on their level of personalization, this paper proposes an evaluation tool grounded on functional benchmarking. Following the steps of benchmarking provided by Basu [3], the building blocks of the evaluation tool are identified. Once the evaluation tool is established, it can be applied to evaluate the level of personalization of an existing PIS and help to identify areas of improvement to reach a higher level of personalization.

3.1 Scope and Partners

The first step in benchmarking involves defining its scope. In the context of personalization, the three core functionalities are designated as the benchmarking scope. Every PIS implemented in practice should have some system behaviors in place that execute the aim of the core functionalities. Given that these core functionalities determine the respective roles of both the passenger and the system in information personalization, they serve as indicators for assessing the personalization level of a PIS. Furthermore, any performance gaps observed in the system behaviors of the core functionalities can provide valuable insights for developers, guiding them toward achieving advanced personalization capabilities.

Second, in selecting a benchmarking partner, this study employs the levels of personalization and their corresponding characteristics. These serve as a foundation for best practices at each level of personalization. The established definitions and characteristics regarding the roles of the passenger and the PIS [2], enable the evaluation tool to conduct a comprehensive gap analysis.

3.2 Performance Measures

When the scope and benchmarking partner are known, it is possible to define the performance measures of benchmarking. In the context of personalization, the system behaviors required to perform the core functionalities serve as relevant performance indicators. According to the levels of personalization [2], the three core functionalities evolve in their aims across different levels, requiring increasingly sophisticated system behaviors to meet these aims. However, the framework on levels of personalization does not identify explicit system behaviour. Therefore, this study needs to identify the system behaviours. To identify the relevant system behaviors, the method of functional decomposition is applied to each core functionality across all personalization levels. This method breaks down each core function into its constituent sub-functions, which represent concrete system actions with defined inputs and outputs [4]. These actions are operationalized through system behaviors. The presence, absence, or partial implementation of these behaviors indicates whether the system is functionally performing as it should. While they do not directly assess user satisfaction or technical efficiency, they provide a structured and replicable way to benchmark what the system does in relation to what it should do at each level of personalization.

3.3 Data Collection and Analysis

Once the performance measures of the evaluation tool are identified, it is possible to collect the data on the performance measures. In the context of personalization, PIS are examined through system documentation, technical specifications, and user manuals to discern the intended functionalities and behaviors of the system. Another method is field testing, which provides empirical evidence of how the system operates in real-world scenarios, and it allows observing system behaviors in practice.

For this short paper, two commercial PIS have been analyzed through document analysis and field testing between March 2025 and May 2025. The two commercial PIS are the mobile application of Stockholm's public transport authority, "Stockholms Lokaltrafik" (SL), and the mobile application of the Dutch train authority, "Nederlandse Spoorwegen" (NS). The two applications are selected based on the accessibility of information regarding the system's functionalities, the diversity in personalization strategies employed, and the ability of the researchers to conduct field testing with the PIS. With the data collected, the PIS's system behavior at each functionality can be compared to the performance measures. The two commercial PIS have been assigned gradient indicators, representing the extent to which the specific system behaviour is implemented. The indicators for implementation are 'not implemented', 'partially implemented' and 'fully implemented'. Attached to the implementation indicators is a scoring tool, which determines a readiness score for every level of personalization. The scores are given as follows: 'not implemented' = 0 points, 'partially implemented' = 1 point, 'fully implemented' = 2 points. The results of the evaluation offer a clear overview of the extent to which the system meets the requirements of each personalization level. These scores facilitate comparative analysis, revealing strengths and weaknesses relative to the PIS and highlighting areas for targeted improvements.

4 Performance Measures per Level of Personalization

To provide a clear overview of the performance measures for each personalization level, this section introduces the sub-functions of the core functionalities. The sub-functions have been derived from core functionality descriptions in the framework of levels of personalization by van Ardenne et al. [2]. Analysis of these sub-functions reveals a consistent logical progression across the core functionalities: first, identification of relevant attributes; second, data acquisition for these attributes; third, data processing and potential labeling; and fourth, application of processed data to personalize travel information. Two sub-functions are optional: fifth, data storage for later use; and sixth, periodic attribute updates to ensure accuracy. The outcomes of these sub-functions per core functionality are summarized in Table 1.

The extent to which passengers and the system are involved in the execution of the sub-functions varies across different levels of personalization. For example, PIS with level 3 personalization are more likely to have automated processes

Table 1. Sub-functions per core functionality of personalization

	Target identification	Situational awareness	Timing of provision
1	Determine which personal preferences and characteristics are required for personalization	Determine which situational attributes are required for personalization	Determine which attributes are required for the provision of the travel information and set conditions for trigger
2	Collect data on the personal preferences and characteristics	Collect data on the situational attributes	Ensure that data on trigger attributes is collected (either as part of data collection of other functionalities or on its own)
3	Process the personal data and add labels to data if needed (with care)	Process the situational data and add labels to data if needed	Process relevant data on trigger and check if trigger conditions are met
4	Apply processed personal data to personalize travel information	Apply processed situational data to personalize travel information	Apply trigger and provide the travel data to the passenger
extra			
5	Stored processed data in a personal profile	X	X
6	Update personal profile with new incoming data	Update situational attributes with new incoming data	X

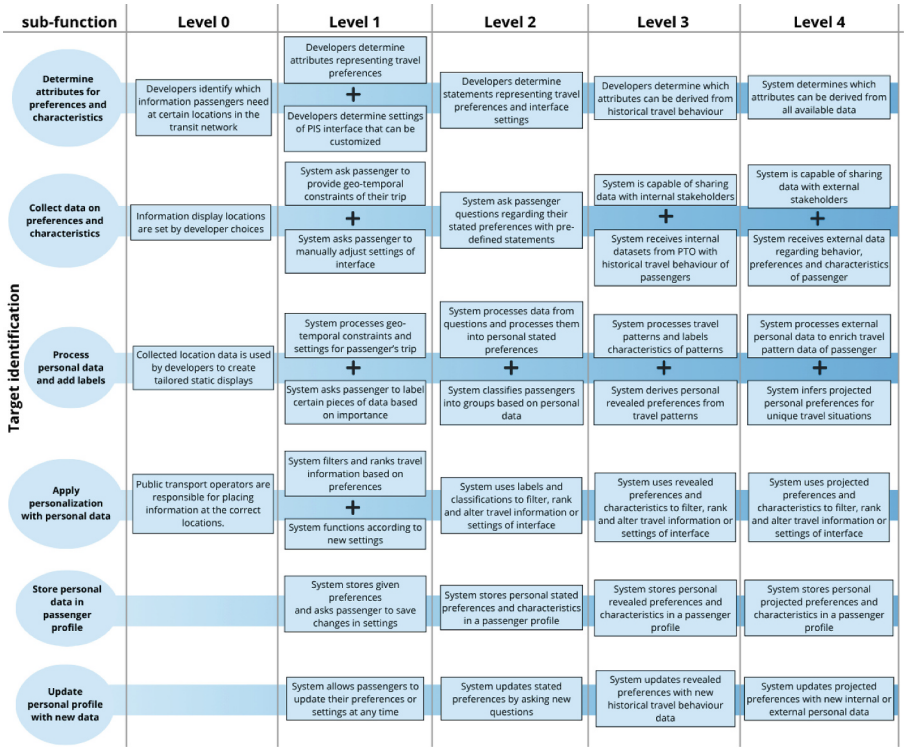

Fig. 1. System behaviour of target identification.

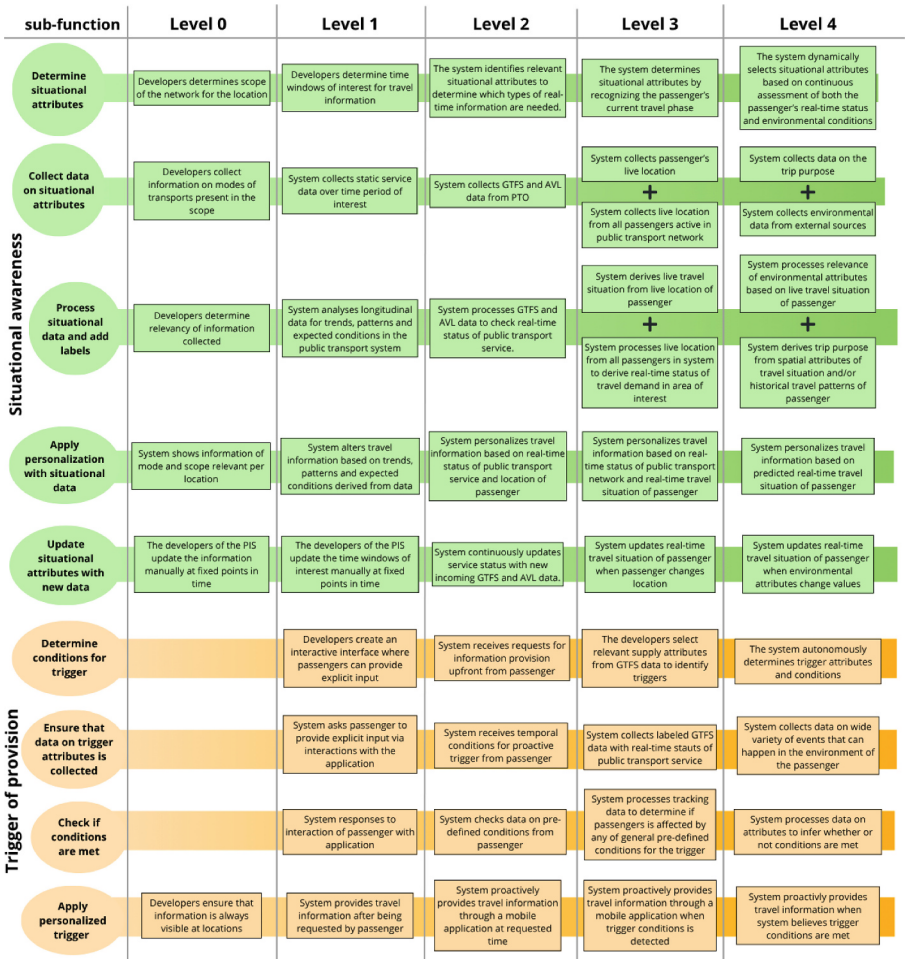

Fig. 2. System behaviour of situational awareness & trigger for provision.

for the data collection on passenger preferences, whereas at lower levels, the sub-functions of target identification rely on explicit passenger input to collect data on the preferences. Moreover, it is possible that the developer of the PIS is involved in the execution of the sub-function. Therefore, it is important to determine the system behaviors in relation to these sub-functions at each level of personalization.

The system behaviors per personalization level are shown in Figs. 1 and 2. The figures illustrate typical system behaviors observed at each level of personalization, organized to indicate the stakeholder initiating each system behavior.

It is important to note that the system behaviors depicted in the figures represent common practices found in both academic literature and commercial PIS. These behaviors should be viewed as operational examples or potential design patterns, rather than strict requirements for achieving a specific level of personalization. The actual implementation will vary depending on the unique context and objectives of each PIS. However, when evaluating a PIS, it is important to consider the initiator and the intended outcome of actions described in the system behaviors. These factors embody the core principles of the personalization levels.

5 Evaluation of Commercial PIS

This section presents the results of applying the evaluation tool to assess the level of personalization offered by the two selected commercial PIS: the mobile applications of SL and NS. For the assessment of personalization in the applications, the system behaviors of the applications are compared with the system behaviors per level of personalization as provided in Figs. 1 and 2.

5.1 Findings for SL Mobile Application

The SL app is a mobile application that allows passengers to plan their travels and buy tickets for the public transport network of Stockholm. With the use of filters, passengers can indicate their preferences for the route recommendations. Moreover, the app provides real-time information on arrival and departure times, as well as cancellations or service details. The results of the evaluation are summarized in Fig. 3.

As shown in Fig. 3, the evaluation reveals that the SL app's system behavior primarily aligns with Level 1 personalization. This is evidenced by the fact that level 1 is the highest level with a full score. Looking at the behaviour of the SL app, it is obvious that target identification largely relies upon explicit data input from the passenger. Passengers actively enter their geo-temporal constraints via a search query. Additionally, the system prompts users to specify travel preferences through the use of filters. These actions correspond with the system behavior associated with Level 1 target identification, wherein the passenger plays an active role in providing and labeling data to facilitate personalization. Furthermore, the system behaviors related to the trigger of provision also demonstrate characteristics indicative of Level 1. The SL app delivers travel information exclusively when the passenger actively engages with the application, such as by sending in a search query.

A notable exception to this pattern is observed within the situational awareness functionality. Here, the SL app showcases the integration of real-time travel information to provide passengers with timely updates on arrival and departure times. The automatic delivery of supply awareness is consistent with Level 2 personalization. This reduced need for passenger participation takes away a burden

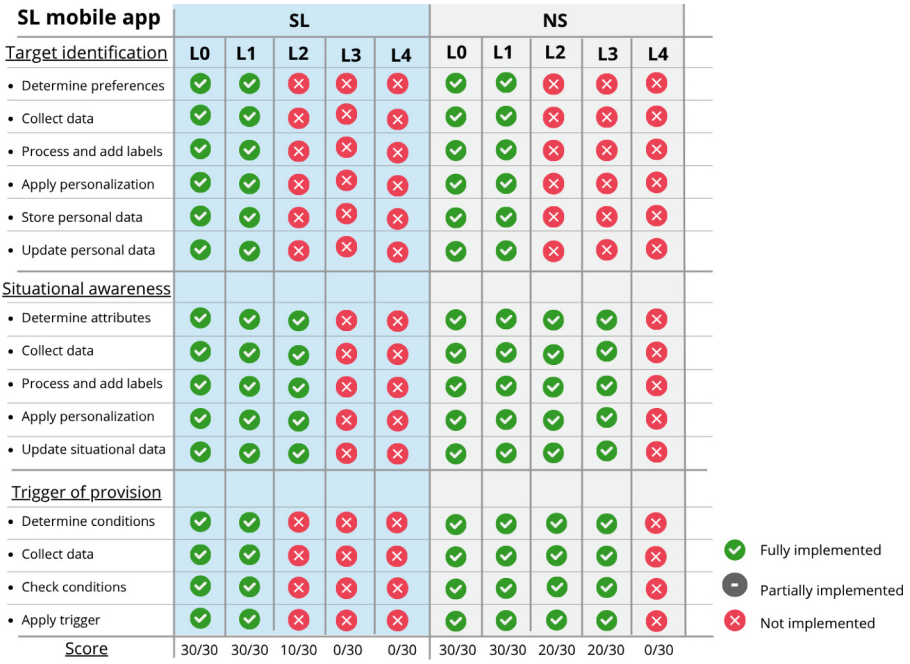

Fig. 3. Evaluation of mobile application of SL& NS app.

on passengers, as the system now assesses vehicle delays, which in turn simplifies the determination of transfer feasibility and timely arrival at destinations.

While the SL app is primarily oriented towards level 1 personalization, opportunities exist to enhance the personalization. Referring to Figs. 1 and 2, future development should focus on implementing stated preferences instead of relying solely on filters. By employing a concise stated-choice questionnaire, the SL app can identify preferences that the passenger is not explicitly aware of, thereby enabling improved personalization options. To enhance user experience, it is crucial to enable passengers to store their stated preferences in a dedicated passenger profile, eliminating the need to complete the stated choice questionnaire for each trip planning session. Moreover, to achieve level 2 personalization, the SL app should let passengers choose when to receive specific information and deliver it proactively, improving user experience and reducing the need to search for updates.

5.2 Findings for NS Mobile Application

The NS app is a mobile application that enables passengers to plan their travels and buy tickets for the national train network of the Netherlands. With the use of filters, passengers can indicate their preferences for the route recommendations. Moreover, the app provides real-time information on arrival and departure times,

as well as cancellations or service details. Unique is the fact that passengers can personalize the notifications provided by the app, to receive updates regarding the status of their planned route. The results of the evaluation are summarized in Fig. 3.

Figure 3 shows that the NS app aligns with level 1 personalization, despite the focus on level 3 personalization for situational awareness and triggers. The NS app provides real-time travel information and keeps track of the passengers' live travel situation to a certain extent. The app displays real-time departure and arrival times of trains, and changes in tracks based on the location of the passenger. Furthermore, the NS app enables users to activate automated push notifications for specific journeys, delivering real-time updates should deviations from the planned itinerary occur. The application leverages its pre-existing capability to process and interpret GTFS and GPS data, utilizing this information as a trigger for providing relevant notifications.

Yet, the NS app is categorized at personalization level 1 due to the system behaviour of target identification, which still relies on explicit data input from passengers regarding a restricted set of geo-temporal constraints and preferences. This system behavior prevents the application from receiving a full score for level 3 personalization. Following the evaluation methodology of Van Ardenne et al. [2], a PIS is classified based on its lowest level of personalization. Therefore, the NS app remains at Level 1, even with the majority of its system behaviors facilitating level 3 personalization.

The evaluation of the NS app indicates that the NS app is close to transitioning to level 3. To make this transition, future development should focus on reducing the passengers' participation in the identification of the preferences of the target. The NS app should invest in models that process revealed preferences from smart card data to uncover preferences for specific travel situations without the need for passenger participation. The implementation of the OV-chipcard system by NS offers a solid foundation for these system behaviours, as it enables the systematic collection and attribution of historical travel behavior for individual passengers.

5.3 Discussion and Conclusion

This study introduces an evaluation method to assess the personalization level of passenger information systems (PIS). The evaluation of two commercial PIS demonstrates that analyzing system behavior can reveal development opportunities for advancing personalization, with a shared finding being the limited automation in the identification of personalization targets. While current literature emphasizes revealed preferences, largely due to the potential of smart card data [6], future research should examine how revealed preference data can be processed and stored in personal passenger profiles to enable more dynamic personalization. Privacy plays an important role here [5], as historical travel data can include sensitive personal information. Additionally, further investigation is needed into the interdependencies among system behaviors across core functionalities and personalization levels. Understanding these dependencies can

clarify which behaviors must be implemented first, particularly when aiming to achieve more complex personalization capabilities. If a desired system behavior relies on several foundational functionalities, those prerequisites must be in place to ensure successful implementation. The lack of a passenger profile, for example, may partly explain why revealed preference-based personalization remains underdeveloped in cases like NS, even though progress has been made in other areas. Future work should focus on mapping these interrelations, particularly those involving data storage, security, and privacy, to support the development of a strategic roadmap for implementing targeted personalization in PIS.

References

1. Vredenborg, M., van Kasteren, A., Masthoff, J.: Personalization in public transport passenger information systems: a systematic review and framework. ACM Comput. Surv. (2025). https://doi.org/10.1145/3721478
2. Van Ardenne, M.T., Cebecauer, M., Cats, O., Ma, Z.: Personalised passenger information systems in public transport: a review and a 5-level personalisation taxonomy. Transp. Rev. **45**(6), 1016–1047 (2025). https://doi.org/10.1080/01441647.2025.2537203
3. Basu, R.: Qualitative Techniques. In: Productivity Press eBooks, pp. 323–372 (2022). https://doi.org/10.4324/9781003268239-18
4. She, J., Belanger, E., Bartels, C.: Evaluating the effectiveness of functional decomposition in early-stage design: development and application of problem space exploration metrics. Res. Eng. Design **35**(3), 311–327 (2024). https://doi.org/10.1007/s00163-024-00434-w
5. Lee, C.-H., Cranage, D.A.: Personalisation–privacy paradox: the effects of personalisation and privacy assurance on customer responses to travel Web sites. Tour. Manage. **32**(5), 987–994 (2010). https://doi.org/10.1016/j.tourman.2010.08.011
6. Cats, O.: Identifying human mobility patterns using smart card data. Transp. Rev. **44**(1), 213–243 (2023). https://doi.org/10.1080/01441647.2023.2251688