TUM SCHOOL OF MANAGEMENT
SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY
TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Information Systems

Evaluating the adoption of the Government as a Platform Analysis Method by practitioners

Efstratios Pahis
03674244
Master in Management and Technology

TUM SCHOOL OF MANAGEMENT

SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Information Systems

Evaluating the adoption of the Government as a Platform Analysis Method by practitioners

Evaluierung der Adaption der Government as a Platform Analyse Methode durch Praktiker

Author:	Efstratios Pahis
Supervisor:	Prof. Dr. Florian Matthes
Advisor:	Dott. Mag. Peter Kuhn
Submission Date:	09.04.2023

I confirm that this master's thesis in information systems is my own work and I have documented all sources and material used.

Munich, 09.04.2023 Efstratios Pahis

Acknowledgments

I would like to especially thank my supervisor Peter Kuhn for his great support throughout my thesis. With your ongoing guidance, dedication, and advice, I had the opportunity to learn a lot from you on a technical as well as personal level. I enjoyed working with you!

Furthermore, I would like to thank Dr. Liudmila Zavolokina for her valuable feedback during my thesis.

I would also like to take the chance to thank Prof. Dr. Florian Matthes. Attending his classes and seminars woke my interest in challenges in business information systems and, finally, motivated me to write my thesis at his chair for Software Engineering for Business Information Systems (sebis). I thank him for the opportunity.

Finally, since this is my last scientific contribution in the foreseeable future and my academic career is approaching its end, I would like to thank also my parents, Dr. Georgios Pahis and Eleni Metaxa. Without them and their unconditional support and dedication, I would not be here writing this thesis. Thank you!

Abstract

Government as a Platform (GaaP) is a recent approach aiming toward the digitalization of the public sector. The core idea of GaaP is to organize the government in terms of an open platform where different public, as well as private entities, contribute collaboratively to the creation of new public services. Harnessing upon the principles of an open platform, governments expect to achieve higher efficiency in the delivery of public services by providing citizens and private companies access to the government's digital infrastructures and, thus, fostering co-innovation. The Government as a Platform Infrastructure Analysis Method (GaaPIAM) is designed to support the transformation towards GaaP by decomposing the existing infrastructure and recomposing it into a platform-oriented one. However, the GaaPIAM has not been extensively evaluated in practice yet. Therefore, the goal of this thesis is to evaluate the GaaPIAM regarding its effectiveness, supportability, and adoption by practitioners in practice. To achieve this, we follow the Design Science Research (DSR) paradigm and develop an evaluation concept for the evaluation of the GaaPIAM in three steps. First, we distill the requirements for the concept. Second, we develop the evaluation concept itself. Finally, we apply the concept in two different workshops with practitioners in order to perform the actual evaluation of the GaaPIAM. This thesis shows that GaaPIAM supports the practitioners to better understand the GaaP approach and, further, identify gaps in their infrastructure and develop actionable proposals for improvements based on the identified gaps. The developed evaluation concept allowed for a precise and rigorous evaluation due to the nine developed criteria aligned with the artifact's goals. Furthermore, its configuration with multiple suitable evaluation methodologies from literature combined with the plethora of evaluation events and modes endorsed the precise and rigorous evaluation of the GaaPIAM. However, although the introduction of "adoption by practitioners" as an evaluation criterion might be useful, utilizing the recommendation-willingness to measure the adoption by practitioners bears drawbacks. From a practical perspective, this master thesis contributes towards the adoption of the GaaPIAM in practice and, therefore, supports governments in their effort to implement GaaP.

Contents

1 Introduction

Government as a Platform (GaaP) is a nascent approach advancing the transformation of the governments' digital infrastructures by following platform-oriented architectures and principles [1, 2, 3, 4]. It envisions the government as an open platform where government and private entities can collaborate and innovate [1, 2]. Thus, similarly to open platform concepts encountered in the private sector, the rough vision of the GaaP approach regards the government as a mere infrastructure provider with governance responsibilities on which multiple entities can create new public services [2, 3, 4]. This approach allows for efficiency gains [2] by limiting the use of resources and "citizen sourcing" [5] while fostering innovation and user-friendliness [1, 2]. Current implementations of the GaaP approach demonstrate its potential and global attractiveness for several governments worldwide, including the UK, Estonia, Italy, India, and Russia [6, 7, 8, 9, 10, 11].

From a technical perspective, Bender and Heine provide a rough blueprint architecture composed of open and modular components for the GaaP approach [12]. At the same time, other research contributes ways to configure and orchestrate these components [4, 8]. However, implementations of the GaaP approach analyzed in the literature indicate a heterogeneous landscape of GaaP architectures and solutions with various technical focuses, often depending on the countries' political, societal, and historic context [5, 7, 13, 14, 15, 16].

Because of the plurality of real-world implementations, implementing the GaaP approach remains challenging in practice [17]. The lack of guidelines and sole case-by-case analysis in the literature intensifies the challenges of governments' endeavors to platformize their digital infrastructures. Therefore, Kuhn et al. propose the Government as a Platform Infrastructure Analysis Method (GaaPIAM) to support governments in transforming their digital infrastructure towards more platform orientation [18]. It primarily supports practitioners in assessing the platform character of the governments' digital infrastructures by promoting the understandability of the GaaP approach, identifying gaps in the current infrastructure required for platformization, and the derivation of design-related actionable proposals [18].

The GaaPIAM is iteratively developed following the Design Science Research (DSR) paradigm. Within the DSR paradigm, the evaluation constitutes a crucial component of the artifact development [19, 20]. However, the GaaPIAM has not yet been extensively evaluated. The previous iterations focused primarily on identifying strengths and weaknesses during the evaluation [21] with the purpose of refining and evolving the GaaPIAM. Consequently, in the current iteration, the revised artifact comprising its previously identified strengths and eliminating its previously identified weaknesses must now be evaluated with a strong focus on its effectiveness, its impact by supporting practitioners, and its adoption-likelihood by the practitioners to rigorously determine its utility. Therefore, this thesis at hand aims at evaluating those objectives and determining the ultimate utility of the GaaPIAM.

1.1 Research Question and Approach

The thesis at hand seeks to evaluate the GaaPIAM according to the iterative DSR paradigm. The evaluation part of the paradigm determines if the artifact does what it was built for, usually along a set of criteria [19, 20]. It essentially "puts the Science into Design Science [Research]."[22].

We focus on evaluating the GaaPIAM regarding the evaluation criteria developed in chapter 5.

The theoretical backbone of this thesis is based on two literature reviews. First, a systematic literature review on GaaP to better understand the context environment of our evaluation and identify related work regarding the artifact. The second literature review concerns DSR Evaluation theory to accumulate an extensive pool of theoretical concepts for evaluating artifacts by utilizing a back-and-forward-based citation search with prominent papers in the field as an anchor point.

The evaluation objective of this thesis is achieved by answering three logically interconnected research questions leading to the final evaluation results of the GaaPIAM.

Research Question 1: What are the criteria for evaluating the GaaPIAM?

The first research question concerns the identification of evaluation criteria relevant to the evaluation of the GaaPIAM. Therefore, we perform an evaluation goal analysis of the GaaPIAM to determine the evaluation criteria [23]. Furthermore, we map the based on the evaluation goal analysis identified evaluation criteria to the GaaPIAM-specific artifact goals. Therefore, we refine the evaluation criteria to better and more precisely capture GaaPIAM's utility. The resulting final set of evaluation criteria comprises, in total, nine criteria. They are organized in three triples, where each triple aims at evaluating one overall evaluation goal for every of the three GaaPIAM artifact goals.

Research Question 2: How can the GaaPIAM be adequately evaluated?

Building upon the insights of the first research question, we then develop an evaluation concept guiding the evaluation of the GaaPIAM. The concept comprises a variety of theoretical concepts and evaluation practices from literature aiming at the rigorous evaluation of our artifact. The developed evaluation concept consists of four stages organized along the nine identified evaluation criteria during the first research question. Hence, we configured every step to best achieve and fit the objective of each criterion. The final configured evaluation concept depicts the concrete content for each evaluation event.

Research Question 3: How can the evaluation concept be applied in practice?

The third research question concerns the practical application of our previously developed evaluation concept and the actual performance of the GaaPIAM evaluation. We applied our evaluation concept in three workshops with two German government entities and conducted

our pre- and post-interviews accordingly. The procedure followed the evaluation concept and, thus, aimed at collecting and analyzing data to evaluate the GaaPIAM's effectiveness, supportability, and adoption by practitioners.

1.2 Thesis Structure

The thesis at hand follows a typical scientific structure. After introducing the topic, research motivation, and research questions, we explain the essential theoretical background and related work. We then briefly demonstrate our research methodology. Subsequently, we structure our result chapters following our three research questions. Therefore, the results of each research question form one complete chapter. Finally, we interpret our results in the discussion section and conclude our work in the conclusion chapter.

2 Foundations

In this chapter, we describe to our best knowledge the current state of the art found in literature, which is fundamental for this thesis. We elaborate on the GaaP approach and DSR Evaluation theory as it serves as the theoretical backbone of our evaluation concept.

2.1 Digital Platforms in the Public Sector

Since the emergence of Information-Communication-Technologies (ICTs), governments have long pursued digitalizing public services. Especially the emergence of the internet in the 1990s intensified the digitalization efforts [24]. Over the years, ICTs evolved, and with them also the approaches and priorities in the digitalization process of the public sector. Those shifts in approaches and priorities can be abstracted into three different phases of the digitalization process, each focusing on different objectives [2].

In the first phase of the digitalization of public services, governments were primarily concerned with mirroring manual paper-based transactions between the government and citizens into digital ones [24]. Exemplified well by government policies such as the amendment to the Paperwork Reduction Act from 1995 in the United States of America [24], which as the name hints, aimed at reducing paper transactions based on digital services [25]. Hence, the digitalization efforts were more technologically-centric, focusing on merely efficiency-driven improvements through digital technologies without disrupting the general operation model of public services [2].

In the second phase, the priorities transitioned from a mere technological-driven view of e-Government to a more citizen-centric one focusing on transforming and reforming the bureaucracy attached to public services [2]. Governments mainly focused on removing redundancies on all levels of public service delivery, from the infrastructures to policies [24].

In the third and current phase, global trends and events such as austerity, the COVID pandemic, and the emergence of new technologies increase the pressure on governments to innovate their public service creation and delivery while downsizing their resource utilization [1, 2]. Moreover, citizens expect more user-friendly public services [2, 8] and service coverage in even the remotest corners of their countries' territory [15]. In response to those challenges, governments increasingly seek the adoption of platform-oriented infrastructures as a potential solution [6, 7, 8, 9, 10, 11]. O'Reilly first framed these adoption efforts of platform-oriented mechanisms and principles as GaaP [1]. Its successful adoption can result in higher efficiency for public service delivery and have innovation-fostering effects through open collaboration and participation with citizens and private entities [1, 2]

2.1.1 Government as a Platform (GaaP)

O'Reilly first coined the approach concerning the platformization of governments' digital infrastructures as GaaP [1]. In his seminal paper, he envisions transferring existing platform concepts and principles from the private sector to the government's public services [1]. Hence, the core idea of GaaP is to organize the government in an open platform where different public and private entities contribute collaboratively to create new public services [1]. The central role of the government in this paradigm is to provide an infrastructure where these entities and the government can build upon, extend, and re-use components to create new public services [1, 2, 4].

The actual configuration and implementation of the GaaP approach entail a series of principles, components, and resources. Although no unified definition [26] and no clear architectural blueprint of GaaP exists, research offers sufficient common aspects detailing the approach.

One crucial aspect is openness, which enables collaboration [2]. It involves open assets, services, and engagement to break down silos of the government's digital system and enhance cooperation among government entities providing public services [3, 4]. This openness is often displayed by shared infrastructures, processes, data, assets, resources, content, and tools [3, 4, 8]. Hence, this openness requires interoperability with other platforms and individual platform components to enable a range of resource sharing [2, 8]. However, this further insinuates a need for orchestration and balancing centralization and decentralization within the government [2, 4, 8] and, subsequently, the need for a suitable choice of top-down or bottom-up approaches[3, 4].

Similarly to the literature about GaaP, no unified view nor blueprint regarding GaaP from the real world exists. However, Jamieson et al. aggregated observations of a GaaP implementation depending on different technical views [27], indicating how GaaP might be approached by governments depending on the adopted views. The first view describes a purely technical approach, which focuses solely on the technical implementation of the platform in the public sector, assuming the technical implementations will eventually impact society [27]. Specifically, this view tends to adopt successful current and past technologies [27]. Therefore, governments adopting a view like this might tend to implement technological aspects akin to GaaP to achieve objectives observed in commercial platforms. For example, Russia approaches GaaP with a purely technical view, neglecting some crucial political and societal implications of GaaP [28]. The second view is more organizational-centric because it implies that technologies related to GaaP are authorized rather than naturally adopted by government bodies [27]. Those decisions to implement technology are often triggered by external events like austerity or changes in citizen behavior like the expectation of 24/7 availability of digital services [2]. Germany can be identified as a country adopting this organizational-centric view as its digitalization efforts of the public sector are mandated and authorized in the form of laws called "Onlinezugangengsgesetz." [29, 30]. The third view combines the purely technical with the organizational-centric view and adds socio-technical aspects [27]. It mainly suggests the emergence of new technology through multiple and reciprocal interactions with different technologies over time instead of the backward,

static views mainly based on past and current experiences with technology or societal and organizational shifts [27]. Governments adopting such a view acknowledge the need for GaaP implementation to have a constant inflow of data to flexibly adjust to user needs and enable reconfiguring and evolving [27]. This highly flexible and evolving property of GaaP can imply the need for integrating and orchestrating multiple platforms, which, again, form a core principle of the GaaP approach [8]. Countries whose implementations of GaaP are considered advanced adopt such a socio-technical view and subsequently build their platform infrastructures with a data-centric, flexible, and mutable focus [6, 7, 8]. Examples of countries adopting such as view while implementing the GaaP approach are the UK, Italy, Estonia, and India [6, 7, 8, 9, 10]. The UK, for example, built an ecosystem of platforms introducing a comprehensive online portal for browsing and accessing all digital public services in one central platform, a digital marketplace, and digital citizen identification services [6]. Also, Italy created several modular platform components, such as a payment service and a digital identification service, to be used interoperably between digital government services [8]. Estonia, like the UK, provides a central online portal for digital public services and a data exchange infrastructure for interoperability and collaboration between government services [7]. Lastly, in India, the government has built a massive and scalable platform-based system to identify and process billions of citizen data to which other digital government services can connect and use [9, 10].

As stated above, there exists no clear blueprint for implementing GaaP in practice. However, Bender and Heine aggregated architectural elements required for the GaaP approach and poured them into a theoretical blueprint depicting GaaP [12]. This blueprint envisions the GaaP approach in the form of a one-stop-shop platform whose primary function is interoperably integrating public services and all required resources [12]. It further entails a platform core providing core functionalities and boundary resources upon which other services can build [12]. Additionally, it offers a marketplace where users can navigate through the existing services on the platform [12]. Lastly, it also should be able to integrate third-party contributions to the platform, ultimately enabling innovation and elevated value delivery to thrive [12]. Platform governance is also crucial to the complete platform ecosystem assuring quality [12].

2.1.2 Implementation Characteristics of the GaaP approach in practice

As described above, there is no unified theoretical definition of GaaP and no practical blueprint of how the GaaP approach might look in practice. However, countries worldwide still digitalize and transform their public service delivery adhering to the GaaP approach. These platform-oriented transformation efforts share some common characteristics regarding the effects on governments' current digital infrastructure, the system-related design decisions taken, and government-stakeholder relations [2, 3, 5, 8]. Additionally, the existent digital infrastructures and organizational-related aspects profoundly influence the government systems' transformation process toward platform orientation [7].

In the following, we group these characteristics along the dimensions (1) platform architecture, (2) platform governance, (3) platform principles, and (4) platform management as they

have been identified as crucial to the GaaP approach [18].

GaaP Architecture Characteristics

Data architecture and open data are one of the literature's most integral and widely mentioned aspects regarding the GaaP approach. Literature describing an advanced and complete approach to GaaP assigns data interoperability, and data exchange a strategic central position and views it as an enabler for a thriving platform ecosystem [27, 31]. The successful utilization of open data in the GaaP approach depends on some preconditions and design decisions. For example, open data must be consistently and correctly versioned to be meaningful and processable for creating better public services and higher value [5, 11, 13, 14].

Data retrieving and the respective API definitions should also be straightforward and easily usable [5, 13]. Therefore, ease of use and simplicity must be integral concepts of a proper platform-oriented architecture [5, 13]. However, this should not be exclusive to the end-user requesting public services but also for other government agencies and third-party contributors [7, 11, 13]. Hence, interfaces must also be easily usable for engineers and designers who integrate different systems and exchange data while implementing the GaaP approach. Hence, API technology and API definitions shall be simple and facilitate access to data [13].

Additionally, for data to be meaningful and meaningfully utilized, it must provide metadata and be linked to each other [5, 13, 14]. However, privacy and security regarding open data for citizens and users are also crucial [3, 7]. Therefore, data ownership should be clearly defined and transparently communicated, and constantly reviewed for adherence to data privacy laws [7]. An organizational and societal impact accompanying the GaaP approach and its open data principle is increasing transparency and democracy [13]. Although this might be considered favorable in many western and democratic societies, countries with more autocratic governments implementing the GaaP approach seem to face multiple challenges with data interoperability, transparency, and exchange [14, 16]. These affected countries often require much more effort to develop data architecture models [14], link data [13, 14, 31], and break up isolated data silos [16].

Moreover, in the case of Estonia, after its independence from the Soviet Union, it did not inherit any digital infrastructure or architecture, allowing it to create a system from scratch without considering integrating any legacy systems [7]. This absence of legacy systems has been shown to facilitate the implementation of the GaaP approach in Estonia [7]. In contrast, other countries with legacy systems, and thus legacy architectures, must consider and integrate them, which adds further complexity while implementing the GaaP approach [11].

GaaP Governance Characteristics

A crucial part of implementing the GaaP approach constitutes integrating systems and services [12], resulting in a system consisting of platform of platforms [8, 12]. This system of platform of platforms inherently causes multiple government agencies and third parties to

join forces and collaborate, amplifying the question of proper role definitions and divisions, especially regarding the distribution of responsibilities and accountabilities [5, 32]. Hence, in their early design decision, governments must define the responsibilities of all actors' roles within the platform-oriented digital infrastructure and determine which party and stakeholder are responsible for what part of the implementation [32]. Simultaneously, they must arrange and agree upon accountability in case of failure or other challenges [5, 32]. Subsequently, governments have to adapt their stakeholder relations and potentially recognize and define new roles resulting from the platformization of their infrastructure [5, 33].

Furthermore, the need for a keystone leadership role among all involved parties is similarly crucial when implementing the GaaP approach. Countries designing and implementing the GaaP approach in a joint venture fashion with multiple stakeholders wish for a broad spectrum of input in a bottom-up manner [9]. Nevertheless, these countries still consider a keystone leadership role assumed by one of the stakeholders critical for successfully orchestrating and compromising all the received inputs and efforts [7, 9, 10]. Therefore, the idea of a dedicated, central entity among equals appears to be necessary [9, 10].

GaaP Principles Characteristics

Another consideration while implementing the GaaP approach is the required openness to feedback and adaptability [2, 3, 7, 13]. Designers and implementors of platform-oriented infrastructures should have a built-in mechanism for receiving feedback from any stakeholders by design and be able to adapt their design decisions and implementations accordingly [2, 3, 7, 13]. This mechanism can range from direct user-experience feedback [13] to elaborate data analytics for feedback analysis purposes [7]. Moreover, open data and easy access to it not only constitute a characteristic of the platform architecture but represent also a fundamental characteristic of platform principles [5, 13]. The exchange and use of open data between the governments' digital infrastructures and other entities are one of the core principles which enable co-participation [2] and allow for collaboration with third parties to develop public services [5, 32, 33].

GaaP Management Characteristics

Platform management represents another important dimension of the GaaP approach, which among others, involves the orchestration and organization of all involved entities [2, 4]. Estonia and UK can be considered spearhead countries in implementing the GaaP approach [6, 7]. They seem to face their challenges more successfully in contrast to Germany, which is a country also pursuing the transformation of its digital infrastructures, in places, towards more platform orientation [17]. Although it is challenging to accurately pinpoint the reasons for the discrepancy based on current literature, a clear distinction between these countries regarding how they organize the transformation of their existing IT systems can still be drawn. On the one hand, the UK and Estonia organize their transformation efforts centrally and follow a straightforward top-down approach [6], which is currently seen as advantageous in regards

to GaaP [7]. On the other hand, Germany is a federal republic comprised of sixteen highly decentralized states with sovereign IT systems and heterogeneous IT providers. However, as other literature has stated, orchestration is one of the essential elements of the GaaP approach [2, 4, 8]. Thus, orchestration is naturally much more challenging in a decentralized and autonomous setting, like in Germany, as opposed to the more centralized, top-down approach in the UK and Estonia [6, 7]. An indication for this observation might be also Germany's recent establishment of the government agency FITKO whose primary goal is to coordinate the transformation processes among all states within the federal republic [17].

2.2 Evaluation of Design Science Research (DSR) Artifacts

DSR's primary goal is the scientific creation of novel solutions for business and domain-specific problems [19]. It differentiates itself substantially from its main counterpart scientific disciplines, such as natural and behavioral sciences, which both focus on the development and verification of theories aiming at the explanation and prediction of naturally occurring phenomena as well as human and organizational behaviors [19, 34]. DSR is also among the scientific disciples the most nascent one [35]. These scientifically created novel solutions are referred to as artifacts as they are created by humans and not occurring naturally [36]. The artifacts' purpose is to solve specific domain problems, often relying upon and applying knowledge and theories developed by their counterpart scientific disciplines in the natural and behavioral sciences [19]. It is hence imperative for an artifact to be innovative and purposefully address a specific domain problem [19, 20]. However, artifacts can range from any tangible and intangible item, such as a piece of software, to a managerial method and are thus practically scopeless [20]. Yet, novel-created artifacts are subject to thorough, scientific evaluation to ensure scientific rigor and its utility [19, 20]. Therefore, the evaluation activity of an artifact is as crucially important as the actual build activity [20].

Thus, DSR Evaluation is one of the core activities in the DSR paradigm [19, 20, 35]. The general purpose of the evaluation activity in DSR concerns the question, "how well does it work?" [20]. Further research refines the scope of the evaluation stage, emphasizing the importance of the artifact's business context during the evaluation and calling for rigorous scientific methods [19]. During the ongoing scientific conversation regarding DSR, research emerged that classifies artifacts also as theory [34, 35, 37]. This classification as a theory implies that artifacts must be stateable as testable, falsifiable propositions, similar to scientific disciplines such as natural and behavioral sciences [34, 35, 37]. Therefore, the evaluation activity encompasses a hypothesis-testing approach to corroborate the artifact's purpose [35, 37].

Nevertheless, a broad consensus exists regarding the instantiation of artifacts in their intended real world context. The instantiation serves as an adequate medium to demonstrate the feasibility of the artifact and evaluate it in its real-world context [19, 20, 34]. This ultimately also enables applying evaluation methodologies from behavioral and natural sciences as the once artificially created artifact becomes a naturally occurring phenomenon in the real world [19].

Evaluation Philosophies and Approaches

Moreover, there exists a vast and heterogeneous amount of proposed requirements and methodologies regarding the concrete approach to DSR evaluation. Often these proposals differ in their philosophical approach toward the evaluation, their evaluation requirements and criteria, and evaluation methodologies. Furthermore, different proposed frameworks offer a multitude of options to configure all of the above.

The two main philosophical views towards DSR evaluation differ in their degree of practicality and theory alike evaluation due to their different view on artifacts [19, 20, 35, 37, 38]. Some research suggests that artifacts are not theory outcomes, while others explicitly regard artifacts as theory outcomes. Hence, the evaluation approaches not regarding artifacts as theory suggest a more practical evaluation focusing more on the business context [20, 39]. In comparison, research that regards artifacts as theory outcomes relies more on a theory-testing approach through formulating testable and falsifiable hypotheses about the artifact's utility [35, 37].

However, later research accepted and incorporated both philosophical views in its work and perceived them as complementary [40, 41].

Evaluation Criteria

Furthermore, a substantial part of DSR Evaluation focuses on choosing appropriate evaluation criteria. These evaluation criteria form the basis of each evaluation as they determine the ultimate goal of the evaluation. March and Smith first proposed a set of in total twelve evaluation criteria for four artifact types [20]. For example, the proposed evaluation criteria for methods as artifact types are ease of use, efficiency, generality, and operationality [20]. Additionally, for instantiations, the proposed evaluation criteria encompass effectiveness, efficiency, and impact on the environment and artifact's users [20]. Further research enhanced this list by adding more quality-related evaluation criteria, such as consistency, accuracy, reliability, fit with the organization, and completeness [19]. Other research introduces the evaluation criteria' importance, accessibility, and suitability to improve the practical relevance of DSR outcomes [42]. On the other hand, research adopting a more theory-based philosophy view artifacts as theory [38]. Thus, the evaluation criteria differ from the above as they are derived and closer related to the more traditional scientific disciplines such as behavioral and natural sciences [38].

Choice of Evaluation Criteria

Choosing the suitable evaluation criteria from the existing plethora of criteria should follow a goal-driven approach [19, 23, 39]. Some research proposes more static evaluation goals that should always aim for utility, quality, and efficacy [19]. Other research proposes choosing more flexible and with the artifact's purpose-aligned evaluation criteria [39, 23]. Hevner et

al. specifically proposes a goal-driven process model for choosing evaluation criteria [23]. One of the process model's objectives is to support the translation of DSR project goals into suitable, goal-aligned evaluation criteria [23]. This process model is structured as an iterative cycle consisting of three phases [23]. At the beginning of the DSR project, rigorous problem space analysis is crucial [23]. Thereby, a strong focus shall lie on analyzing the problem space context, like the spacio-temporal setting and the environment [23]. Furthermore, the evaluation's goals must be defined, ranging from utilitarian, safety, interaction and communication, cognitive and aesthetic, innovation, and evolution goals [23]. In this initial phase, suitable evaluation criteria must be selected based on the goals and the context [23].

In the second phase, located in the solution space, the artifact must be built and formatively evaluated utilizing the selected evaluation criteria to assess and refine the artifacts [23]. In the third phase, the artifact must be summatively evaluated against the DSR goals using the selected evaluation criteria again [23]. The artifact can be delivered if the evaluation results satisfy the defined goals [23]. Whereas not, the artifact must repeat the cycle to be further refined [23].

Evaluation Methods (regarding Data Collection)

After the determination of evaluation goals as well as the evaluation criteria, the selection of the appropriate evaluation methodology is imminent. For that purpose, research proposes many methodologies, some also well-known in other scientific disciplines and some others more specific to the practical nature of the DSR paradigm. Prominent research distilled five distinct evaluation methodology types for evaluating artifacts [19]. The first evaluation methodology type is observational, which involves case and field studies [19]. The second evaluation methodology is analytical, which analyzes different artifact qualities and involves static, architecture, optimization, and dynamic analysis [19]. The third is experimental, which aims to grant more control of confounding and moderating factors and involves controlled experiments and simulation with artificial data [19]. The fourth evaluation methodology type is testing which involves functional testing and structural testing [19]. Lastly, descriptive is the last evaluation methodology type and involves an informed argument using the existing knowledge base and scenarios demonstrating the artifact's utility [19]. Further research complements this list of evaluation methodologies, such as laboratory research, field inquiries, surveys, action research, prototyping, and expert evaluation [43, 44].

Evaluation Frameworks

Furthermore, research proposes a plethora of evaluation frameworks to support the apt configuration of evaluation goals, criteria, and methods. The proposed frameworks differ specifically in goal, focus, and scope. Table 2.1 displays an excerpt of selected prominent DSR evaluation frameworks.

Authors	Philosophical Approach	High Level Characteristics
Hevner et al. [19]	More practical evaluation approach	General guidelines for evaluation; Loosely interlinked set of evaluation criteria and methodologies
Pfeiffer et al. [43]	More practical evaluation approach	Mapping of artifact type, evaluation criteria, and evaluation approaches
Peffer et al. [44]	More practical evaluation approach	Mapping of artifact type, evaluation criteria, and evaluation methodologies
Sonnenberg et al. [45]	More practical evaluation approach	Evaluation patterns guiding ex-ante and ex-post evaluation
Venable et al. [22]	More practical evaluation approach	Evaluation methodologies selection matrix based on ex-ante and ex-post evaluation formative and summative evaluation
Venable et al. [46]	More practical evaluation approach	Evaluation strategies guiding evaluation
Venable et al. [37]	More theoretical evaluation approach	Utility hypothesis formulation
Gregor [35]	More theoretical evaluation approach	Hypothesis based evaluation with blueprint formulation
Aier et al. [38]	More theoretical evaluation approach	Criteria configuration for artifacts seen as design theories
Bucher et al. [40]	Practical and theoretical evaluation approach	Selection matrix for evaluation methodologies depending on research gap or real-world gap
Cleven et al. [41]	Practical and theoretical evaluation approach	Complete evaluation configuration matrix comprising multiple attributes

Table 2.1: Overview over selected DSR Evaluation frameworks in literature

For example, Hevner et al. propose a set of guidelines that generally guide through the complete DSR process, including the evaluation of an artifact [19]. Additionally, they are offering evaluation criteria and methodologies but without comprehensively interlinking them [19]. Some other prominent frameworks offer a selection matrix for evaluation methods based on the artifact's ultimate evaluation goal, which can be the evaluation against a research gap or the real world [40]. Similarly, other frameworks exist that offer a selection matrix based on a mapping of artifact type, evaluation criteria, and evaluation approach [43]. Peffer et al.'s framework also aims at mapping evaluation criteria and evaluation methodologies to artifact types [44]. However, they further justify the mapping results via a statistical analysis regarding the kind of evaluation methodologies DSR research papers used for their artifacts [44]. Additionally, other frameworks also acknowledge the temporal dimension of an evaluation, which can occur ex-ante and ex-post [22, 45, 46]. The ex-ante evaluation of an artifact is formative [22, 45, 46]. Therefore, it affects the artifact in its prototype phase, whereas the ex-post evaluation is more summative and evaluates the actual instantiation of an artifact [22, 45, 46].

Furthermore, these frameworks introduce an artificial and naturalistic dimension to capture the evaluation context better [22, 46]. A naturalistic evaluation context occurs when multiple diverse stakeholders exist, the artifacts have a socio-technical nature, the evaluation focuses on the artifact's effectiveness, or the artifact shall be evaluated close to the real world context [22]. On the other hand, an artificial evaluation context occurs when few similar stakeholders exist, the artifact has a purely technical nature, the evaluation focuses on the artifact's efficacy, and firm control over confounding and moderating variables is desired [22]. Based on the above frameworks, Venable et al. further refined and extended their contributions and molded them into the FEDS Framework [46]. The FEDS framework offers four evaluation strategies based on the evaluation context [46]. The first is quick and simple, which shall be used

for low, social, and technical risks [46]. The second one is human risk and effectiveness, which is suitable for social or user-oriented artifacts, relatively cheap to evaluate with real users and context, and if the goal of the evaluation is to determine the sustainable utility in the real world [46]. The third one is technical risk and efficacy, which is more suitable for technical-oriented artifacts and where the evaluation with real users in the actual setting might be too expensive [46]. The focus lies here in demonstrating the efficacy of an artifact, hence, the utility arising solely from the artifact itself [46]. The last strategy is purely technical artifact and shall be utilized when the artifact is purely technical and does not involve any social aspects [46]. Additionally, such a strategy may be chosen if the artifact's intended use is in the future and is not designed for the present [46].

Other Frameworks offer an extensive matrix capturing the full DSR evaluation dimensions and properties [41]. These frameworks consider the methodological approach, the artifact focus, artifact type, epistemology, function, the actual method, the object, the ontology, perspective, the position reference point, and finally, time [41].

Lastly, some frameworks follow a more theory-based philosophical approach, viewing artifacts as theory outcomes [35, 37, 38]. By classifying artifacts as theory, these frameworks propose formulating utility theories similar to hypotheses from other scientific disciplines [35, 37]. These must be testable and falsifiable and aim to demonstrate an artifact's utility against the artifact's non-existence, basically the status quo [35, 37]. However, these approaches should complement other more practical approaches like those described above, not excluding them [37].

3 Related Work

In the following section, we will introduce the related theoretical background regarding the iteratively created GaaPIAM, which is this thesis's primary subject of evaluation.

3.1 GaaP Analysis Approaches

GaaP literature has proposed several concepts aiming at describing platform-related properties, classifying GaaP-oriented implementations, and providing blueprints for major required components in a GaaP implementation. Table 3.1 displays an overview of the different GaaP analysis approaches and their conceptual focus.

Analysis Tool	Reference	Dimensions	Procedure	Goal
GaaPIAM	Kuhn, et al. [18]	Platform architecture Platform governance Platform openness Platform management	Yes	"Provide a comprehensive framework of what to consider when applying GaaP in practice, identify gaps of the infrastructure with regard to GaaP, [and] create actionable proposals for the improvement of the infrastructure toward GaaP" [18]
Platform Assessment Framework	Brown et al. [6]	Organizational form Market dynamic Architectural structure	No	Auditing platform initiatives aiming "to provide consistency of thinking in GaaP initiative" [6]
Platform of Platforms Framework	Cordella and Paletti [8]	Organization Architecture	No	"Help public administration to deliver public value better" in GaaP initiatives by introducing the platform of platforms architecture [8]
Platform Architecture Blueprint	Bender and Heine [12]	Ecosystem Platform Core, Boundary Resources Governance	No	Identify platform elements in the public sector and integrate them into a public platform concept [12]
Socio-Technical Analysis Framework	Jamieson et al. [27]	Technical Analysis View Social Analysis View Socio-technical Analysis View	No	Assess how the government's designs and implementations of the GaaP approach manifest based on view [27]
Open Government Data Platform Classification Framework	Danneels et al. [31]	Cognitivist View Connectionist View Autopoietic View	No	Assess how the government's designs and implementations of platforms manifest based on view [31]
Open Governance Framework	Millard [3]	Open Assets Open Engagement Open Services	No	Offer a high-level configuration framework for the GaaP approach [3]

Table 3.1: Identified GaaP Analysis approaches in literature

Kuhn et al.: GaaPIAM

The first GaaP tool is Kuhn et al.'s GaaPIAM [18]. This method offers a procedure to assess the government's digital infrastructure regarding its platform orientation. This procedure

must be performed along four dimensions regarding the government's digital infrastructure: the platform architecture, platform roles, platform openness, and platform governance and management [18]. Detailed information and analysis of the GaaPIAM reside in the section 3.2 below.

Brown et al.: Platform Assessment Framework

The second concept is the Platform Assessment Framework from Brown et al. [6]. They propose a framework for auditing platform initiatives aiming "to provide consistency of thinking in GaaP initiative." [6]. The proposed framework offers not only a purely technical dimension for the assessment but also incorporates the organizational form, market dynamic, and architectural structure [6]. Each dimension provides concepts and features from Information Systems literature for assessing the GaaP approach [6]. They applied the framework to the UK's government's digital infrastructure, which follows the GaaP approach [6]. However, it does not provide a specific procedure or method to apply the framework [6].

Cordella and Paletti: Platform of Platforms Framework

The third concept is from Cordella and Paletti and aims to "help public administration to deliver public value better" [8]. The framework focuses on the orchestration as it suggests configuring GaaP implementations as a "Platform of Platforms" [8]. Therefore, it consists of an organizational dimension as well as an architectural dimension [8]. The framework has been instantiated to Italian GaaP initiatives [8]. However, also this framework does not provide any procedure to apply the framework [8].

Bender and Heine: Platform Architecture Blueprint

The fourth concept originates from Bender and Heine, which aims to identify platform elements in the public sector and integrate them into a public platform concept [12]. The platform concept is described as a one-stop-shop infrastructure functioning as an integration mechanism for third-party and government services [12]. It encompasses crucial elements for GaaP-oriented systems such as the platform ecosystem, architecture, and governance derived from literature [12]. Additionally, they apply the public platform concept in different European countries and the UK to examine its characteristics [12]. However, it does not provide any procedure for applying it to support GaaP initiatives [12].

Jamieson et al.: Socio-Technical Analysis Framework

The fifth concept is from Jamieson et al. and ties socio-technical views regarding technology to observations from practical GaaP initiatives [27]. Thus, it provides a framework for assessing how the government's designs and implementations of the GaaP approach may look based on the adopted view [27]. This concept has been applied to the GaaP initiatives

in the UK [27]. Nevertheless, like the other concepts, it does not provide any procedure for applying it [27].

Danneels et. al: Open Government Data Platform Classification Framework

Danneels et. al first defined Open Government Data platforms, whose core idea is similar to the GaaP approach but with a stronger emphasis on open government data [31]. Nevertheless, the framework uses GaaP literature without explicitly utilizing the expression GaaP [31]. It is the sixth concept, and besides defining Open Government Data platforms, they offer a framework to assess how different adopted views regarding platforms influence the platforms' instantiation along five dimensions [31].

Millard: Open Governance Framework

Finally, the seventh concept originates from Millard and offers a blueprint configuration framework, called the open governance framework, applicable to any GaaP initiative [3, 4]. The framework consists of three configuration areas composing the GaaP approach: open engagement, open services, and open services [3, 4]. Each configuration area offers multiple configuration options and considerations [3, 4]. However, the framework does not offer a concrete procedure to configure a GaaP initiative, nor has it been applied to any initiative in practice [3, 4].

3.2 The GaaP Infrastructure Analysis Method (GaaPIAM)

Although there are successful implementations of the GaaP approach in the real world [6, 7, 8], governments still face challenges and barriers in their transformation process [17]. A potential cause might be the lack of analysis tools and procedures supporting governments in assessing their current infrastructure for its infrastructures' platform character. Thus, Kuhn et al. proposed the GaaPIAM to support the transformation of the government's digital infrastructure toward a more platform-oriented one [18]. The method's main objective is to provide a concrete procedure to analyze governments' digital infrastructures regarding their platform character by

"(a) provide a comprehensive framework of what to consider when applying GaaP in practice, (b) identify gaps of the infrastructure with regard to GaaP, [and] (c) create actionable proposals for the improvement of the infrastructure toward GaaP." [18]

The theoretical background below is based on Kuhn et al.'s paper [18] if not cited otherwise.

GaaPIAM Dimensions

GaaPIAM consists of four dimensions, along which its procedure is later organized.

The first dimension is platform architecture which deals with the infrastructure components and their interactions comprising a GaaP architecture. These infrastructure components are the platform core, platform boundary resources, and the ecosystem. The primary purpose of this dimension is to identify crucial components of the infrastructure that can be later mapped to platform architecture.

The second dimension concerns platform governance, which encompasses the stakeholders of a platform infrastructure and their respective roles. These platform roles can be platform owners, platform complementors, and users. The primary purpose of this dimension is to identify the stakeholders and roles of digital infrastructures and subsequently further identify absent or double-assigned roles by mapping them to the platform ones.

The third dimension is platform principles, which concern use cases related to interactions between complementors, users, and the platform infrastructure. It particularly tracks the alignment with the platform principles: platform openness, participation, and co-creation. The primary purpose of this dimension is to identify use cases entailing interactions with the platform infrastructure to asses them along the three above mentioned platform principles. This assessment further supports the identification of weaknesses in the platform's openness.

Lastly, the fourth dimension deals with platform management. The primary focus in this dimension lies in managing platform assets and activities, facilitating the orchestration of platform infrastructure. The primary purpose is to identify gaps and inconsistencies in terms of platform governance.

GaaPIAM Procedure

The procedure proposed by the GaaPIAM composes two main activities along the four above introduced dimensions. The two main activities can be further broken down into four sub-tasks.

The first main activity is decomposition and consists of the sub-tasks: status quo and infrastructure decomposition.

The first sub-task, status quo, concerns identifying major elements in the current infrastructure along the four dimensions. Thus, in the platform architecture dimension, identify the significant infrastructure components; in the platform governance dimension, identify the prominent actors; in the platform principles dimension, identify the major usage scenarios; and in the platform management dimension, identify the infrastructure's owner's primary activities.

The second sub-task, infrastructure decomposition, concerns classifying the identified major elements from the first sub-task along the four dimensions. Thus, in the platform architecture dimension, classify the identified infrastructure components according to their importance for the core infrastructure use cases; in the platform governance dimension, classify actors according to the identified actors' roles; in the platform principles dimension, classify according to the identified usage scenarios' infrastructure services; and in the platform management dimension, classify according to the identified activities' stakeholders.

Completing these above explained sub-tasks also concludes the decomposition activity, and

the newly classified elements now require reconfiguration to interpret them in terms of their platform character. This reconfiguration occurs in the second main activity: the recomposition. The recomposition activity, like the decomposition activity, also consists of two sub-tasks.

The first sub-task, platform interpretation, of the recomposition activity concerns mapping the identified and classified elements from the decomposition activity along the four dimensions. Thus, in the platform architecture dimension, map the identified and classified infrastructure components to the platform core, boundary resources, and ecosystem; in the platform governance dimension, map the identified and classified roles to the platform owner, complementors, and user; in the platform principles dimension, map identified and classified infrastructure services to the platform principles: openness, participation, and co-creation; In the platform management dimension, map the identified and classified activities to orchestration and facilitation activities: facilitation, tool provision, and asset management.

The second sub-task of the recomposition activity is the assessment and concerns inferring recommendations and actionable proposals from the previous mapping activity along the four dimensions. Thus, the inferred recommendations are bound to the dimensions of platform architecture, platform governance, and platform principles and platform management.

The assessment sub-task concludes the recomposition activity and completes the whole procedure of the GaaPIAM. The gained insights can now support practitioners to "consolidate the identified gaps and develop actionable proposals" [18].

The below Figure 3.1 visualizes the procedure of the GaaPIAM.

	Platform architecture	Platform governance	Platform principles	Platform management
THEORY	Learn and discuss central concepts of GaaP and its dimensions			
ANALYSIS				
1. Status quo	Identify the major components of the infrastructure	Identify the major actors of the infrastructure	Identify the major usage scenarios for the users and complementors	Identify the activities of the infrastructure owner(s)
2. Infrastructure decomposition	Classify the components to the infrastructure core use cases	Classify the actors by their role	Classify the scenarios by infrastructure services	Classify the activities by stakeholder
3. Platform interpretation	Map the components to the platform core, boundary resources, and ecosystem	Map the actors to the roles of platform owner, complementor, and user	Map the infrastructure services to openness, participation, and co-creation	Map the activities to facilitation, tool provision, and asset management
4. Assessment	Infer proposals for the improvement of the platform	Infer proposals for the improvement of the platform	Infer proposals for the improvement of the platform	Infer proposals for the improvement of the platform
CONSOLIDATION	Consolidate the identified gaps and develop actionable proposals inspired by examples from other countries			

Figure 3.1: Steps of the GaaPIAM, according to Kuhn, et al. [18]

GaaPIAM Research Methodology

The research methodology of the GaaPIAM is based on the Design Science Research Paradigm.

Specifically, it follows a six-step research approach aligned with Peffers et al. [21, 44]. This six-step research approach was performed in two iterations. The first iteration aimed at building the artifact and evaluating its understandability [21]. The second iteration aimed to refine the artifact based on the first evaluation and evaluate its applicability and usefulness [21].

GaaPIAM Evaluation

The latest GaaPIAM evaluation was conducted in form of workshops with three German government agencies. Each agency brought a specific real world infrastructure of interest to apply the GaaPIAM. The data was collected through observations during the workshops and follow-up interviews conducted directly after the workshops [21]. The evaluation criteria chosen are primarily based on Sonnenberg and Borcke's criteria of effectiveness, user-friendliness, efficiency, completeness, level of detail, internal consistency, generality, and specificity [21, 45]. Every criterion was evaluated using a Likert scale question and two open questions [21]. The GaaPIAM evaluation results show that each criterion's objectives are generally met [21]. However, the evaluation did not focus extensively on the actual effectiveness of the GaaPIAM [21]. It additionally does not demonstrate if its application supports practitioners to impact the infrastructure's transformation towards platform orientation or the adoption of practitioners.

4 Methodology

The general objective of the thesis at hand is the evaluation of the GaaPIAM. The GaaPIAM is an artifact created through the building and evaluating activity within the DSR paradigm [19]. Hence, this evaluation is part of the overall DSR cycle of the GaaPIAM. Therefore, to structure our evaluation of the GaaPIAM, we align with Hevner's evaluation guidelines [19] and his three cycles framework [47]. Since our evaluation is part of a DSR Project and does not constitute a standalone project, we position our evaluation concept in the relevance cycle of Hevner's three DSR cycles [47]. The final outcome will be an evaluation concept applicable to the GaaPIAM, which again will be applied during GaaPIAM's instantiation.

Thus, as part of the relevance cycle of the GaaPIAM design, we analyze the business and domain environment to develop evaluation criteria aligned with Hevner's proposed goal-driven approach for specifying evaluation criteria based on our evaluation goals [23]. The domain environment, in our case, depends on the instantiation environment where the GaaPIAM will be applied. A particular focus lies on the challenges and opportunities of this environment as it will determine the evaluation concept's final design and, thus, the actual evaluation of the GaaPIAM. Additionally, we conduct a systematic literature review regarding GaaP to further enhance our understanding of the evaluation environment. For the actual evaluation criteria development, we follow three steps aligned with Hevner's framework [23]. In the first step, we aggregate and summarize evaluation criteria from the literature while defining our evaluation goals. In the second step, we then select suitable evaluation criteria corresponding to our evaluation goals. Finally, we map the chosen evaluation criteria to the artifact's main goals and, thus, generate a set of criteria aligned with our evaluation goals and the artifact's goals.

The below Figure 4.1 displays the process of developing our evaluation criteria using the Business Model Processing Model 2.0 technique.

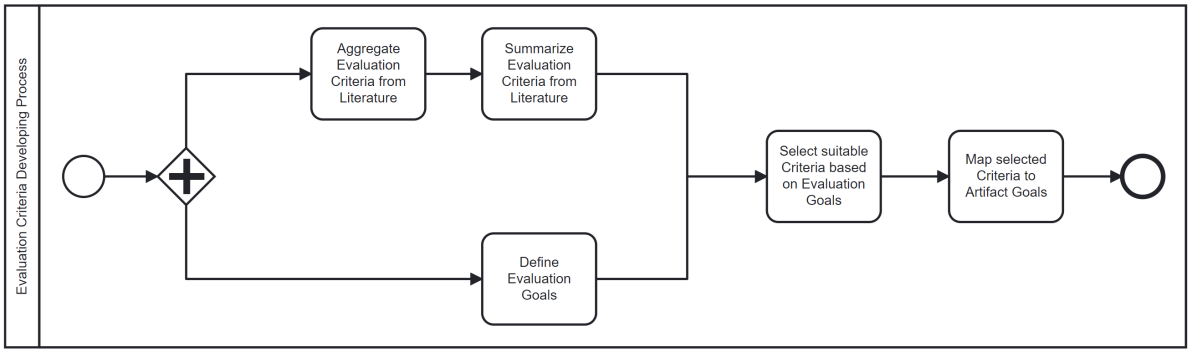

Figure 4.1: Overview of the complete evaluation criteria development process

Subsequently, based on the derived evaluation criteria, we further analyze the existent DSR Evaluation theory to choose suitable evaluation frameworks and methodologies for our evaluation concept. Thus, we conduct a back-and-forward citation-based literature review with prominent research papers from the respective field as anchor points.

We design the evaluation concept iteratively, combining the knowledge gained from the literature review with constant feedback from the researchers who created the GaaPIAM. Later, the concept is applied to the instantiation of the GaaPIAM.

The concrete steps for that purpose are unfolded in the following. First, we analyze the evaluation context to adapt our evaluation criteria accordingly [23]. The context analysis particularly focuses on the challenges regarding the GaaPIAM instantiation environment. Subsequently, we find suitable solutions to the identified context challenges, which we ultimately use to adapt the evaluation criteria. Afterward, for each adapted evaluation criterion, we choose an adequate evaluation methodology according to the state of the art in the literature, a suitable data collection procedure, and a suitable evaluation setting regarding the evaluation timing and execution medium. Consequently, this represents the skeleton of the evaluation concept, which we then adequately configure based on each evaluation criterion.

Figure 4.2 displays the complete evaluation concept configuration process using the Business Model Processing Model 2.0 technique. It additionally highlights the iterative procedure that must be performed for each evaluation criterion, which in our case, accumulates to a total of nine times for our nine developed evaluation criteria.

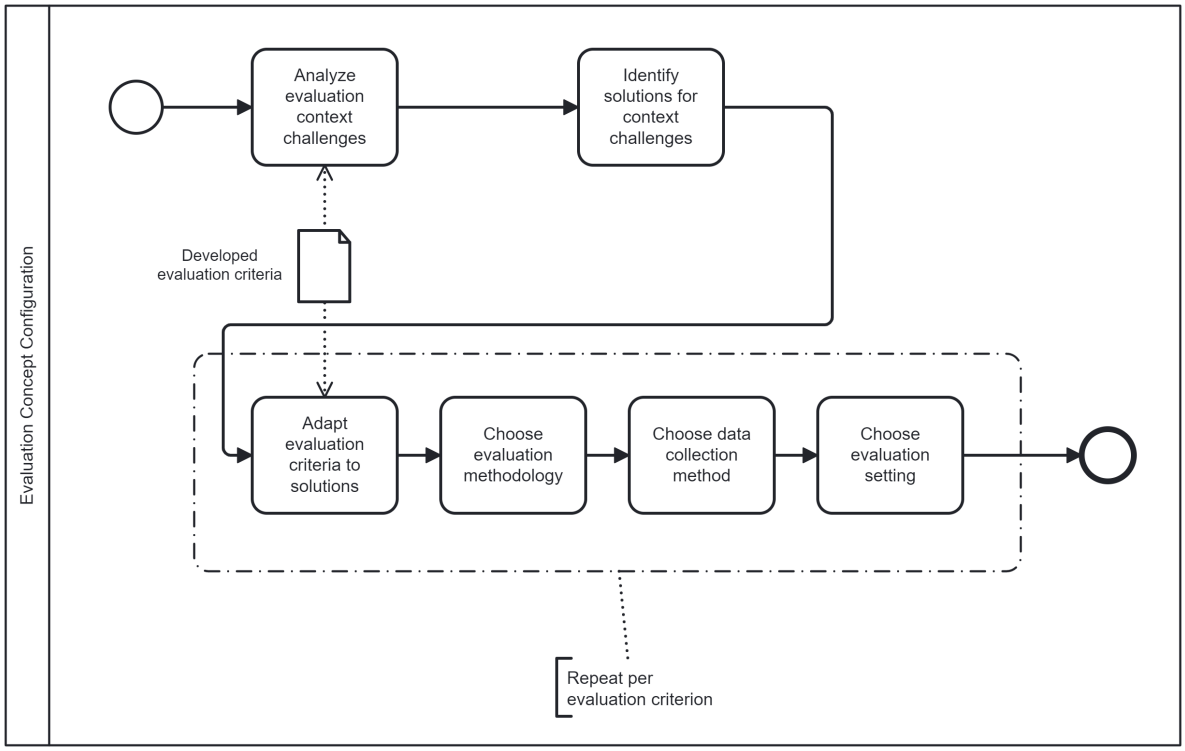

Figure 4.2: Overview of the complete evaluation concept configuration process

4.1 Data Collection

As mentioned above, for our research, we conducted two major literature reviews to create the evaluation concept for the GaaPIAM. We then apply the evaluation concept to the instantiation of the GaaPIAM.

The first systematic literature review regards the topic of GaaP. We utilized Scopus's scientific database and entered the search string "Government as a Platform," which can be contained in the title, abstract, or as a keyword. This search yielded 35 papers, of which seven were not freely available, and one was written in a foreign language. Subsequently, we enhanced the list with five more papers identified through backward-and-forward citations based on the previously compiled literature of 35 papers. However, four papers we consider irrelevant to our research after reading them completely. Hence, in total, 36 papers were taken into consideration.

The second literature review regards the DSR Evaluation theory. We conducted it purely based on the back-and-forward citation technique with the prominent papers from Hevner et al., March and Smith, Gregor, and Venable et al. as an anchor point [19, 20, 35, 46]. Our search yielded 27 papers, of which 20 we consider relevant to designing our evaluation concept. Nevertheless, from those 20, eight papers comprise the actual core of the evaluation concept [20, 22, 23, 34, 35, 37, 46, 47]. Furthermore, we transferred two more papers from the

behavioral sciences discipline to complement the evaluation concept's core [48, 49].

We applied the evaluation concept and, hence, conducted the evaluation of the GaaPIAM during its instantiation in two workshops with two German government organizations. From a pure methodology-relevant perspective, the evaluation concept comprises four principal data collection events: semi-structured pre-interviews, transcriptions and observations during the workshops, survey directly after the workshop, and semi-structured post-interviews. We conducted the pre-interviews directly before the workshops on the same day with an approximate duration of thirty minutes. The main content-wise goals are to identify the interviewees' role in their organization, their prior knowledge of GaaP-related topics, and the current challenges they face in transforming their infrastructure towards more platform orientation. We conducted these semi-structured pre-interviews with all the practitioners of each workshop. We based them exclusively on open questions. We then recorded and transcribed the interviews and then inductively codified them [50]. In the workshops, we apply the GaaPIAM with the practitioners on their own real world digital infrastructures. During the workshops, we gather data by observing how the method supports the practitioners regarding their understanding of GaaP, identifying gaps in their infrastructure, and inferring actionable proposals for improvement regarding their infrastructures. Like the pre-interviews, the post-interviews are semi-structured and conducted with the workshops' practitioners. However, we perform the post-interviews with a temporal distance of approximately three to four weeks, practically enabling capturing how the generated insights from the application of the GaaPIAM in the workshops might have supported the practitioners impacting their infrastructure in that time period. The duration of the post-interviews is approximately forty minutes long.

We conducted the survey with the practitioners directly after the workshop on the same day in the presence of all participants, thus, practitioners and researchers. It consisted of three statements, which the practitioners assess based on their agreement or disagreement. Hence, it resembles a Likert scale-based questionnaire [51]. However, we did not evaluate or reveal the results during the survey to the practitioners to mitigate social desirability distortion bias originating from the face-to-face workshop mode [52]. Therefore, we collected the data digitally and anonymously.

Subsequently, we further posed a ten-point Likert scale-based question regarding the practitioners' willingness to recommend the GaaPIAM for each evaluation criterion. This methodology resembles the Net Promoter Score by Reichheld [48, 49].

Finally, during the workshops, we gathered data through observations and persisted them through field notes. We aligned with current guidelines regarding field notes [53]. Hence, we captured contextual setting information that can support our interpretations and evaluation-relevant data from a subjective and objective nature [53]. For example, we took field notes about the practitioners' engagement with the GaaPIAM and practitioners' insights gained through applying the GaaPIAM. Furthermore, we transcribed the complete workshops, which had an average duration of 3.5 to 4.5 hours.

The below Figure 4.3 displays the chronology of the data collection events paired with the respective data collection methods and time spans.

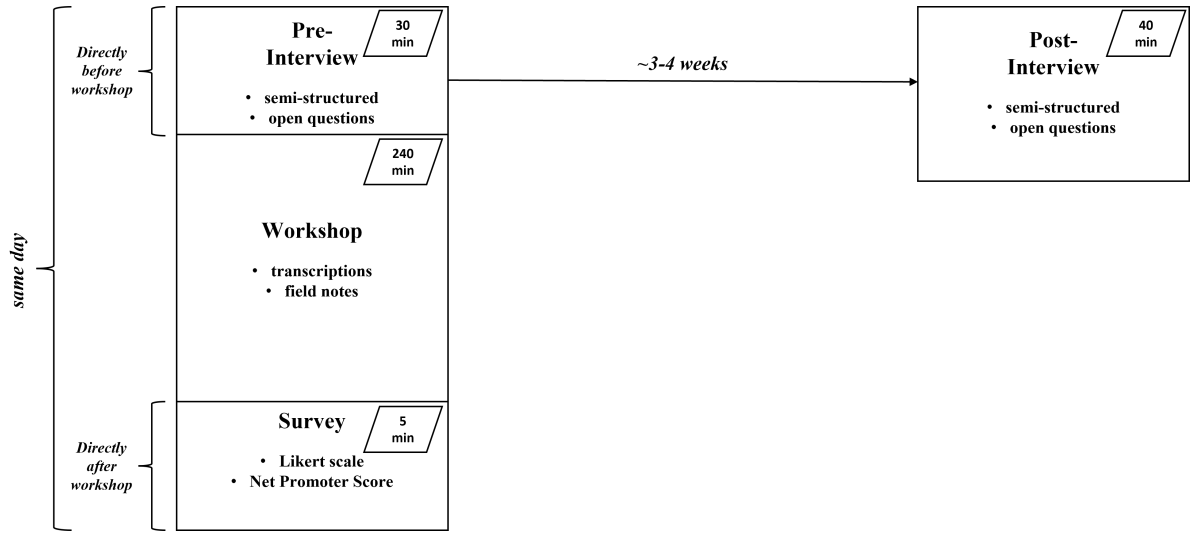

Figure 4.3: Chronological overview of the data collection events paired with the respective data collection methods and time spans

4.2 Data Analysis

Regarding the analysis of the semi-structured pre- as well as post-interviews, we use a mixture of inductive and deductive coding techniques [50]. The inductive coding strategy is mainly utilized in analyzing the challenges-related data of the pre-interview. However, we then utilize these inductively generated codes in a deductive manner to analyze the post-interviews. This analysis regards how the generated insights from the workshops could support the practitioners to impact their infrastructures since the application of GaaPIAM. Hence, the firstly inductively generated codes serve as the baseline for measuring and comparing GaaPIAM's supportability potential better in regard to the practitioners' faced challenges.

Regarding the workshop analysis, we followed a similar approach to the semi-structured interviews. We used workshop transcriptions to codify the data according to our evaluation concept deductively [50]. We additionally utilized the field notes to enrich the workshop data analysis.

Although we collected quantitative data from the practitioners through the Likert scale questions in the survey, the sample size is insufficient to allow for any statistical analysis due to violations of the central limit theorem [54]. Therefore, we analyze the obtained data qualitatively combined with two context-related open questions in the post-interview, consequently ensuring a degree of validity and reliability.

For analyzing the data generated from Net Promoter Score questions, which also constitute a part of the survey, we followed Reicheld's proposed ranges [48, 49]. Therefore, a score between nine to ten depicts the willingness to recommend [48, 49]. In contrast, a score between seven and eight hints at a neutral stance, and with any score below six, we can

conclude that practitioners will likely actively not recommend the GaaP Method [48, 49].

5 Evaluation Criteria

This chapter concerns the development of criteria required for evaluating the GaaPIAM. The developed evaluation criteria will serve as the basis for the overall evaluation concept.

In general, the evaluation of an artifact serves the purpose of determining how well an artifact performs [20]. However, the evaluation criteria further specify what needs to be evaluated to measure the well-working of an artifact. Thus, they are an integral part of the evaluation.

Consequently, the first step in developing an evaluation concept for the GaaPIAM is to determine a suitable set of evaluation criteria.

5.1 Evaluation Goals and Criteria

According to Hevner et al.'s goal-driven approach to developing evaluation criteria, a crucial part is determining the overall evaluation goals for an artifact [23]. To specify the goals, we first analyze the previous instantiation of the GaaPIAM and its evaluation.

In the previous instantiation, the applied evaluation concept followed a broad scope [21]. Thus, the primary focus did lay on generating an extensive set of strengths and weaknesses to refine the GaaPIAM further [21]. Consequently, the utilized evaluation criteria were broadly defined and tried to capture many artifact aspects [21]. In total, eight evaluation criteria from the literature were chosen [21]. They ranged from generality and user-friendliness to completeness and effectiveness [21]. Although this clearly helped refine and significantly improve the GaaPIAM, crucial aspects of the Method were left unevaluated or not thoroughly examined.

The first aspect that was not thoroughly examined is the degree of effectiveness of the GaaPIAM. Though the previous evaluation included effectiveness as an evaluation criterion [21], it was not evaluated extensively and precisely enough. The chosen data collection procedure was based on a single Likert scale question followed by two open questions examining the general effectiveness of the artifact immediately after the instantiation [21]. However, this procedure holds certain drawbacks. First, the researchers asked the Likert scale question directly in a face-to-face mode. This questioning mode is prone to social desirability bias which might lead to better-than-actual results [52]. Secondly, the data collection occurred solely directly after the workshop [18]. Therefore, practitioners were not given the opportunity to thoroughly reflect on the GaaPIAM, meaning that any response can only be a mere indication of its effectiveness based on first impressions. Finally, the lack of specificity in how the question was asked to the practitioners does not allow them to distinguish nor attribute its effectiveness to one of the three GaaPIAM artifact goals.

Hence, we formulate the first evaluation goal as follows:

(1) Evaluate the GAPIAM's degree of effectiveness on the government's digital infrastructure

The second aspect that was not examined in the previous iteration [21] is GaaPIAM's extent of impact on its environment, hence, the government's digital infrastructure. The previous evaluation ignored this aspect [21]. However, we regard it as an essential aspect for evaluating the GaaPIAM and is also emphasized by literature [20]. Furthermore, we refine the scope of impact by tying it to the ability to support practitioners as it resembles one of the prime objectives of the GaaPIAM [18].

Therefore, we include the impact on the infrastructure by supporting the practitioners as one of our evaluation goals and formulate it as follows:

(2) Evaluate the GAPIAM's extent of impact on the government's digital infrastructure by supporting the practitioners

Finally, the third aspect that was not examined in the previous iteration [21] is the practitioners' adoption of the GaaPIAM. The previous evaluation ignored this aspect and did not evaluate if practitioners were willing to utilize the method in their work. However, we consider evaluating the adoption of the GaaPIAM by practitioners critical. Henceforth, we aim to examine the practitioners' sustainable and long-term willingness to use the Method to support the transformation of the government's digital infrastructures towards more platform orientation.

Therefore, we include the adoption of the GaaPIAM by practitioners as one of our evaluation goals and formulate it as follows:

(3) Evaluate the GaaPIAM's adoption by the practitioners

While formulating suitable evaluation goals, we further aggregate evaluation criteria from the literature and summarize them based on their similarity. The following Table 5.1 displays the aggregated evaluation criteria from the literature. However, since the GaaPIAM represents a method as an artifact outcome, we include only evaluation criteria that are applicable to methods and instantiations according to the literature.

[20]	[19]	[43]	[38]	[55]	[56]
Completeness Ease of use Effectiveness Efficiency Generality Impact on the environment and on the artifact's users Operationality	Utility Quality Efficacy Completeness Consistency Accuracy Reliability Usability Fit with the organization	Appropriateness Completeness Consistency Functionality Usability Reliability Performance Supportability	Utility Internal consistency External consistency Broad purpose and scope Simplicity Fruitfulness of new research findings	Validity Utility Quality Efficacy	Accesability Importance Novelty and insightfulness Actability and guidance Effectiveness

Table 5.1: List of identified evaluation criteria in literature

From the above evaluation criteria, we eliminate the ones mentioned multiple times and summarize the closely related ones to narrow down the possible candidates for evaluation criteria. Hence, after the summarization of the 39 identified criteria, 18 remain. The following Table 5.2 displays the 18 summarized evaluation criteria.

Aggregated and summarized Evaluation Criteria
Effectiveness
Efficiency
Impact on the environment and on the artifact's users
Operationality
Quality
Efficacy
Consistency
Accuracy
Reliability
Usability
Appropriateness
Supportability
Validity
Accessibility
Importance
Novelty
Simplicity
Usefulness

Table 5.2: Aggregated and summarized evaluation criteria

5.2 Evaluation Criteria Selection

After formulating the evaluation goals and identifying evaluation criteria from the literature, we then perform a selection of evaluation criteria to use in our evaluation concept based on the defined evaluation goals and the identified criteria [23]. For that purpose, we assess which criteria can best express and capture the evaluation goals. Table 5.3 displays the final assessment and, thus, a pre-selection of evaluation criteria.

Evaluate GAPIAM's degree of effectiveness on infrastructure	Evaluate the GAPIAM's extent of impact on infrastructure by supporting practitioners	Evaluate the adoption of GaaPIAM by practitioners
(Degree of) Effectiveness	(Extent of) Impact on the environment and on the artifact's users (Extent of) Supportability	Usability Accessibility Simplicity Usefulness

Table 5.3: Selected evaluation criteria based on evaluation goals and evaluation criteria from literature

For evaluating the degree of the effectiveness of the artifact, the effectiveness criterion follows exactly that purpose [20], thus, making it the most suitable choice.

For the evaluation goal regarding the impact of the artifact on the infrastructure by supporting the practitioners, we identified two suitable evaluation criteria. The first one is the "Impact on the environment and on the artifact's users" [20] and the second one is "Supportability" [43]. Regarding the latter, we do not mean supportability in the sense of an artifact's maintainability. We rather view supportability as a criterion capturing the ability to support someone achieving a certain objective. Subsequently, following the above distinction, we argue that supportability incorporates, to a significant extent, impact. This is due to the fact that in order for the artifact to support the practitioners, it must also impact them. Additionally, impact as a criterion on its own has a broad scope which might influence the precision of the evaluation. In contrast, the supportability criterion better captures the condition from the evaluation goal by associating the impact to the support provided to the practitioners and, thus, narrows and specifies the evaluation scope.

Regarding the final goal, "Evaluate the adoption of the GaaPIAM by the practitioners," multiple suitable evaluation criteria can be identified and attributed. For example, if practitioners perceive the GaaPIAM as simple, useful, and operational, it might indicate and lead to their adoption of the Method. However, these would only be indications, and reasoning for a causal relationship can not be rigorously established. Therefore, we introduce "adoption by practitioners" as a new evaluation criterion, which has not yet been introduced by literature.

We describe "adoption by practitioners" as the sustainable and recurring use of an artifact by practitioners. Thus, it aims to evaluate an artifact's overall utility by capturing if practitioners utilize it over time. Therefore, it inherently implies that for the evaluation, there is an adequate long evaluation period available to observe the artifact over time in the real world.

Hence, Table 5.4 displays the final selection of evaluation criteria.

Evaluate GAPIAM's degree of effectiveness on infrastructure	Evaluate the GAPIAM's extent of impact on infrastructure by supporting practitioners	Evaluate the adoption of GaaPIAM by practitioners
(Degree of) Effectiveness	(Extent of) Supportability	Adoption by practitioners

Table 5.4: Final set of selected evaluation criteria

5.3 Evaluation Criteria Mapping on Artifact Goals

The GaaPIAM comprises three different sub-goals as introduced in chapter 3 in section 3.2. The first is providing a comprehensive framework of what to consider when applying GaaP in practice and thus increasing the understandability of the GaaP approach [18]. The second is identifying gaps in the infrastructure required for a platform-oriented transformation [18].

Lastly, developing concrete actionable proposals to transform the infrastructure towards more platform orientation constitutes GaaPIAM's third sub-goal [18]. In order to perform a more precise evaluation and obtain detailed results, we further map the three previously developed evaluation criteria to the above GaaPIAM goals. Therefore, we can better differentiate how each aspect of the artifact performs and infer specific, more adequate evaluation results. After the mapping, a total of nine evaluation criteria are distilled. The below Figure 5.1 displays the mapping results and the final evaluation criteria.

Artifact Goals / Evaluation Criteria	**Increase understandability of GaaP**	**Identifying gaps in infrastructure**	**Developing actionable proposals for infrastructure improvement**
Degree of Effectiveness	*Degree of effectiveness in increasing the understandability of GaaP*	*Degree of effectiveness in identifying gaps in the infrastructure*	*Degree of effectiveness in developing actionable proposals for improving infrastructure*
Extent of Supportability	*Extent of support for practitioners to impact their infrastructure due to the increased understandability of GaaP*	*Extent of support for practitioners to impact their infrastructure due to the identification of gaps in infrastructure*	*Extent of support for practitioners to impact their infrastructure due to the development of actionable proposals for improving infrastructure*
Adoption by practitioners	*Adoption of GaaPIAM due to the increased understandability of GaaP*	*Adoption of GaaPIAM due to the identification of gaps in infrastructure*	*Adoption of GaaPIAM due to the development of actionable proposals for improving infrastructure*

Figure 5.1: Mapping of artifact goals to evaluation criteria

6 Evaluation Concept

This chapter concerns the development of the evaluation concept, which will be utilized during the GaaPIAM instantiation. It builds upon the findings from the previous chapter 5. Accordingly, the previously derived evaluation criteria are the very foundation of this evaluation concept.

The configuration of the evaluation concept is designed as a multi-step process, which we outline in chapter 4.

6.1 Evaluation Criteria Adaptation based on Evaluation Context

The first step in configuring the evaluation concept is the adaptation of the evaluation criteria to the evaluation context, which also aligns with current literature [23]. Thus, to identify evaluation context-specific properties, we examine the evaluation criteria with a focus on possible challenges. This examination resulted in three main identified context challenges.

Figure 6.1 displays the examination results and the identified solutions to overcome evaluation context challenges.

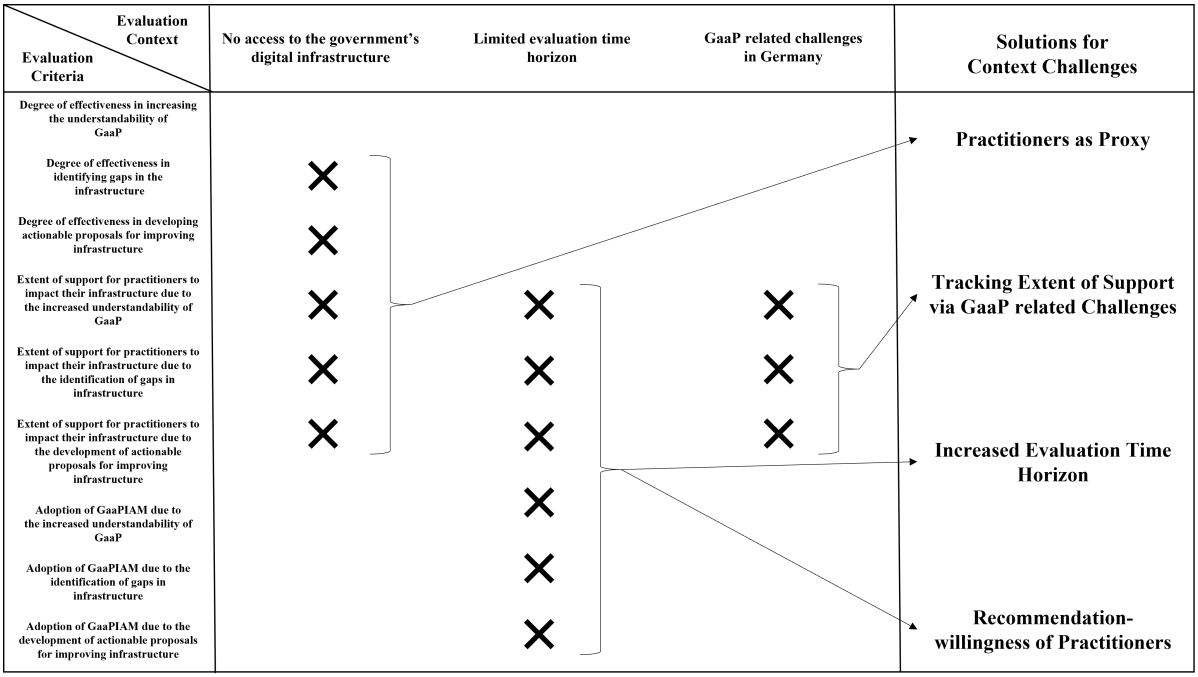

Figure 6.1: Evaluation context analysis and identified solutions to context-related challenges

Context Challenge 1: No Access to Infrastructure

The first context challenge concerns the infrastructure. Multiple evaluation criteria aim at evaluating the degree of effectiveness and extent of support for practitioners to impact their existing infrastructure. However, obtaining access to the existing infrastructure of government systems is highly unlikely and often subject to national security concerns. Additionally, causally attributing any changes regarding the government's infrastructure to the application of the GaaPIAM is highly speculative due to many confounding and moderating variables introducing bias.

A possible solution to this might be using the practitioners as a proxy, who apply the GaaPIAM, instead of utilizing the infrastructure itself as the evaluation subject. The underlying evaluation assumption is that practitioners applying the GaaPIAM will use the insights to transform the government's digital infrastructure they are responsible for. Moreover, accessing the practitioners is much more feasible and simultaneously cost-effective.

Therefore, we adapt all evaluation criteria affected by this challenge and aim to evaluate the GaaPIAM with practitioners as a proxy instead of the actual infrastructure.

Context Challenge 2: Limited Evaluation Time-Horizon

The second context challenge concerns the evaluation time horizon. Multiple evaluation criteria aim at evaluating the extent of supportability and the adoption of the Method by the practitioners. However, capturing the extent of supportability, which we defined as the ability to support practitioners to impact their existing infrastructure, inherently requires a time period with at least two temporally separate evaluation events. This is mainly due to the fact that in order to support the practitioners, there must be an impact on the practitioners observable. Therefore, it requires an evaluation baseline that captures the status quo and an additional evaluation to enable the comparison between the status quo and changes due to the application of the GaaPIAM. These two evaluation events should occur separately before and after the instantiation. The former serves as the baseline and the latter for comparing the baseline to the GaaPIAM's application effects on the extent of supportability. Moreover, the criteria evaluating the adoption of the GaaPIAM are similarly bound to the same temporal hurdle. This is due to the fact that the adoption is a process consummated through sustainable and re-occurring utilization of an artifact over time. Thus, it requires an extensive evaluation time horizon too.

A possible and pragmatic solution to this might be the extension of the evaluation horizon to the maximum the project allows. Secondly, regarding the evaluation criteria for capturing the artifact's adoption, evaluating the recommendation-willingness of the practitioners might be a possible solution. The recommendation-willingness might serve as a suitable measure to capture satisfaction, loyalty, and engagement [48, 49] with the GaaPIAM, which we argue increases the likelihood of adopting the Method over time if given.

Therefore, we adapt all evaluation criteria affected by this challenge and aim to evaluate

the GaaPIAM by increasing the evaluation time horizon as much as the project allows, which is three to four weeks in our case since the first instantiation. Secondly, we also use the practitioners' recommendation willingness as a proxy to evaluate the Method's adoption.

Context Challenge 3: GaaP in Germany

The third context challenge concerns how GaaP is perceived and approached in the German public sector. Kuhn et al. identified ample barriers existing in the German public sector regarding GaaP and platformization of digital infrastructures in general [17]. These challenges range from actually grasping the GaaP approach and communicating it to stakeholders to more technical-related challenges like defining a platform's scope from a technical perspective [17].

At first glance, these challenges the practitioners face when confronted with the GaaP approach do not immediately pose any challenges to the actual evaluation. However, they can be utilized to better capture the impact of the GaaPIAM and, thus, support the evaluation of supportability-related evaluation criteria. Hence, the central idea is to utilize challenges as the subject of interest to track the extent of supportability before and after the instantiation of the GaaPIAM.

Therefore, we adapt all evaluation criteria affected by this challenge and aim to evaluate the GaaPIAM's supportability according to the practitioners' challenges regarding GaaP and the platformization of infrastructure in general.

6.2 Evaluation Concept Configuration

As mentioned above, we utilize the context-adapted evaluation criteria as a starting point for configuring our evaluation concept. However, due to the similarity of the configurations, which only substantially differ between the effectiveness-, supportability-, and adoption-related criteria, we summarize and cluster for the sake of clarity and simplicity the nine evaluation criteria into three substantially differing criteria: effectiveness, supportability, and adoption. Therefore, when referred to effectiveness, it comprises the criteria: (1) Degree of effectiveness in increasing the understandability of GaaP, (2) Degree of effectiveness in identifying gaps in the infrastructure, and (3) Degree of effectiveness in developing actionable proposals for improving infrastructure. Furthermore, when referred to supportability, it comprises the criteria: (1) Extent of support for practitioners to impact their infrastructure due to the increased understandability of GaaP, (2) Extent of support for practitioners to impact their infrastructure due to the identification of gaps in infrastructure, and (3) Extent of support for practitioners to impact their infrastructure due to the development of actionable proposals for improving infrastructure. Finally, when referred to adoption, it comprises the criteria: (1) Adoption of GaaPIAM due to the increased understandability of GaaP, (2) Adoption of GaaPIAM due to the identification of gaps in infrastructure's architecture, and (3) Adoption of GaaPIAM due to the development of actionable proposals for improving infrastructure.

Nevertheless, there still exist some differences also within the clusters. We will address these differences and highlight them accordingly if they occur.

Figure 6.2 displays a high-level overview of the complete evaluation concept configuration matrix skeleton.

Evaluation Criteria		Effectiveness-related evaluation criteria	Supportability-related evaluation criteria	Adoption-related evaluation criteria
Evaluation Context		Practitioners as a proxy	Practitioners as a proxy & increased evaluation time horizon	Practitioners ' recommendation willingness
Evaluation Methodology		Testable, falsifiable Utility-Hypothesis	Ex-post, summative & naturalistic Evaluation based on FEDS	Recommendation based evaluation
Evaluation Data Collection		One Likert scale & two open questions	Semi-structured interview & field notes	Net Promoter Score
Evaluation Settings	Evaluation Timing	Directly after workshop & after workshop with time difference	Directly before workshop & after workshop	Directly after workshop & after workshop
	Evaluation Format	Survey & post-interview	Pre-interview, workshop observations & post-interview	Survey

Figure 6.2: High-level overview of evaluation concept configuration matrix

Corroboration of Evaluation Assumptions

Before performing the actual evaluation of the GaaPIAM along the three criteria, we must verify that our underlying assumptions for the evaluation can be corroborated. Particularly, this affects the context solution from section 6.1 regarding utilizing practitioners with a significant impact on the infrastructure as proxies for the actual infrastructure. To adequately determine the evaluation suitability of practitioners participating in the instantiation of the GaaPIAM, we collect information about their roles and responsibilities in their organization. In this context, we furthermore question them about their motivation to participate and their opinions regarding the need for additional tools to support them in transforming their infrastructure. Consequently, we can further corroborate the relevance of the GaaPIAM for the practitioners [47].

The respective data collection occurs within the pre-interview directly before the GaaPIAM's instantiation.

Figure 6.3 displays the final configured data collection questionnaire to corroborate the evaluation assumptions.

Evaluation Criteria	Evaluation Timing	Data Collection Goal	Question
Non-Specific	Directly before the instantiation	Practitioners' relevant background information	What are your role and primary tasks in your organization?
		Practitioners' familiarity with the GaaP approach and associated support tool	How familiar are you with the Government as a Platform approach?
			What was your motivation to participate in this workshop?
			To what extent does your organization utilize the Government as a Platform approach to transforming its infrastructure?
			What kind of tools do you use to support the transformation of your infrastructure?
			What more tools would you wish to have for transforming the transformation of your infrastructure?

Figure 6.3: Data collection questionnaire for corroborating the evaluation concept assumptions

Effectiveness concerning Evaluation Criteria

The first cluster of evaluation criteria concerns effectiveness-related criteria. As already stated above, for effectiveness-related criteria, we adapt to the evaluation context and use the practitioners as a proxy for the evaluation instead of the actual infrastructure.

In the second step of the evaluation concept configuration, we choose an adequate evaluation methodology. Two main streams of evaluation methodologies can be distinguished in the literature. One stream follows a more practical approach, usually based on frameworks that offer practical configuration options [19, 20, 22, 43, 45, 44, 46]. On the other hand, the second stream follows a more theoretical approach that builds upon the formulation of testable, falsifiable hypotheses [34, 35, 37, 38], also referred to as utility hypothesis [37]. These utility hypotheses are especially suitable for evaluating an artifact's utility, hence also its effectiveness [37]. Furthermore, more nascent literature endorses using hypothesis-based evaluation methods for evaluating effectiveness [56]. We, therefore, choose the utility hypothesis as the base evaluation methodology for evaluating the effectiveness-related evaluation criteria. Thus, we must formulate and pour our criteria into the utility hypothesis. Gregor proposes to formulate the utility hypothesis by applying the hypothesis-development technique well known from the natural and behavioral sciences [35]. The main requirement is that the formulated utility hypotheses are testable and falsifiable [35, 37]. Hence, we formulate and pour the effectiveness-related evaluation criteria into the following utility hypotheses:

(1) Utility Hypothesis: The GaaPIAM will better increase the understandability of the GaaP approach to the practitioners than without it

(2) Utility Hypothesis: The GaaPIAM will better enable practitioners to identify gaps in their infrastructure's architecture than without it

(3) Utility Hypothesis: The GaaPIAM will better enable practitioners to develop actionable proposals for improving their infrastructure than without it

In the third step of the evaluation configuration, we select adequate data collection methods for the developed utility hypothesis. Livari et al. propose using a practitioner-centric approach when evaluating an artifact [56]. Furthermore, to evaluate the effectiveness, they propose formulating statements capturing the hypotheses' objectives and asking them to the practitioners in a Likert scale fashion within a questionnaire [56]. Therefore, aligning with the current literature, we choose to evaluate the effectiveness-related evaluation criteria using Likert scale-based statements derived from the hypotheses [51, 56]. Furthermore, in addition to the practitioners' evaluation on the Likert scale, we ask two open questions to enhance our understanding of their evaluation following the mixed-method research paradigm [57]. The purpose of each open question is to understand how effective the underlying hypothesis goal is or what is lacking to achieve that goal. We complement the collected data with field observations during the instantiation of the artifact.

Figure 6.4 displays the finally configured data collection questionnaire to evaluate each effectiveness-related evaluation criterion.

Evaluation Criteria	Hypothesis	Mode	Question
Degree of effectiveness in increasing the understandability of GaaP	The GaaPIAM will better increase the understandability of the GaaP approach to the practitioners than without it	Likert	The workshop helped me to better understand the Government as a Platform approach.
		Open	How did the workshop help you to better understand the Government as a Platform approach?
		Open	Did any aspects of the Government as a Platform approach remain unclear to you?
Degree of effectiveness in identifying gaps in the infrastructure	The GaaPIAM will better enable practitioners to identify gaps in their infrastructure's architecture than without it	Likert	The workshop helped me to better identify optimization potentials in my infrastructure.
		Open	What optimization potentials could you identify in your infrastructure?
		Open	What more optimization potentials did you hope to identify in your infrastructure?
Degree of effectiveness in developing actionable proposals for improving infrastructure	The GaaPIAM will better enable practitioners to develop actionable proposals for improving their infrastructure than without it	Likert	The workshop helped me to better develop improvements in my infrastructure.
		Open	What improvements could you develop in your infrastructure?
		Open	What more improvements did you hope to develop in your infrastructure?

Figure 6.4: Data collection questionnaire for effectiveness-related evaluation criteria

In the last step, an adequate evaluation setting must be configured. Particularly the timing of the evaluation data collection events and the evaluation data collection medium must be configured accordingly. Regarding the evaluation timing for effectiveness-related evaluation criteria, we collect the Likert scale data from the practitioners directly after the instantiation of the GaaPIAM. Therefore, we collect data that captures the very first impressions from the practitioners while the artifact is mentally very present to them. However, the open questions related to the same evaluation criterion we ask after the instantiation with a significant time

difference, which we aim to be around three to four weeks. Following this approach, we strive to collect data after we give the practitioners the time to evaluate their first impressions and process the gained insights. This approach also enables us to evaluate whether the first impressions are corroborated or overturned, ultimately making the evaluation more rigorous and robust.

We provide the Likert scale questions to the practitioners via a digital, anonymous survey. For the two open questions, we conduct a semi-structured post-instantiation interview.

Supportability concerning Evaluation Criteria

The second cluster of evaluation criteria concerns supportability-related criteria. As already stated above, for supportability-related criteria, multiple solutions for the evaluation context challenges have been identified. Therefore, we adapt the evaluation criteria by using the practitioners as a proxy for the evaluation instead of the actual infrastructure, extending the evaluation horizon to a maximum, and ultimately utilizing GaaP-related challenges to track the supportability.

In the second step of the evaluation concept configuration, we choose an adequate evaluation methodology. From the abovementioned two main streams of evaluation methodologies, we choose a more practical approach. First, we align with the current literature and apply the FEDS framework to identify an adequate strategy for evaluating the supportability-related criteria [46]. The FEDS framework offers four strategies to choose from, as already introduced in the foundation chapter 2 in section 2.2. The chosen strategy then allows for inferring adequate evaluation methodologies. Based on our evaluation project context and setting, we choose to follow the Human Risk and Effectiveness strategy because it is relatively cheap in our project to evaluate with real users in their real context [46]. Additionally, regarding the supportability-related evaluation criteria, we further aim to evaluate the continuous benefits and utility that exists over the long run [46]. The other three strategies do not fit our evaluation project. The Human Risk and Effectiveness strategy proposes a formative evaluation for building an artifact before it actually is instantiated, thus, evaluating ex-ante [46]. However, this part can be considered fulfilled through the previous evaluation iteration of the GaaPIAM leading to improvements of the artifact [18, 21]. Hence, we focus on the more summative, ex-post, and naturalistic evaluation proposed by the Human Risk and Effectiveness strategy [46].

To find suitable data collection methods for a summative, ex-post evaluation and, thus, configure the third step of the evaluation concept, we utilize Venables et al.'s Comprehensive Framework for Evaluation in Design Science [22]. This framework offers an extensive matrix for choosing the suitable data collection method based on the chosen evaluation strategy [46]. For summative, ex-post, and naturalistic evaluations, the framework proposes to choose from case studies, focus groups, participant observations, ethnographies, phenomenologies, or surveys as possible data collection methods [22]. Therefore, for evaluating GaaPIAM, we choose a mixture of participant observations and qualitative surveys. The qualitative surveys' objective is to capture the GaaPIAM's supportability based on the practitioners'

faced challenges. To supplement and corroborate our survey evaluation findings, we observe the practitioners during the GaaPIAM instantiation and take relevant field notes aligned with guidelines from literature [53]. We particularly focus on the practitioners' engagement with the artifact during the instantiation, solutions generated through its application as well as limitations occurring during its application.

Regarding the last evaluation concept configuration, the evaluation setting, we adjust the evaluation timing according to the identified context challenges or, rather, the developed context challenge solutions. Hence, we conduct two temporal distinct evaluation events for the supportability-related evaluation criteria. The first occurs before the actual artifact instantiation with the goal of gathering data about their current challenges regarding the transformation of their infrastructure. We conduct this first evaluation event directly before the actual instantiation on the same day. The second evaluation event is dated after the actual artifact instantiation with the goal of gathering data about how the practitioners' generated insights during the instantiation supported them in facing their challenges. We conduct this evaluation with a temporal distance from the first evaluation of around three to four weeks to allow the practitioners to utilize the gained insights and potentially already impact their respective infrastructures in some way.

However, due to the fact that is not possible to rigorously distinguish if the practitioners have been supported due to an increase in understandability of GaaP, identification of gaps in their infrastructure, or development of actionable proposals for improvement, we summarize the data collection-related questions for all supportability-related evaluation criteria. Hence, we do not distinguish the open questions according to each evaluation criterion.

We perform the survey in the form of a semi-structured post-instantiation interview.

Figure 6.5 displays the final configured data collection survey to evaluate each supportability-related evaluation criterion taking the above described temporal settings into account.

Evaluation Criteria	Evaluation Timing	Data Collection Goal	Question
Extent of support for practitioners to impact their infrastructure due to the increased understandability of GaaP & Extent of support for practitioners to impact their infrastructure due to the identification of gaps in infrastructure & Extent of support for practitioners to impact their infrastructure due to the development of actionable proposals for improving infrastructure	Directly before the instantiation	Practitioners' current challenges regarding the transformation of their infrastructure	What kind of challenges do you face transforming your infrastructure?
			How did you tackle these challenges so far?
	After the instantiation	Practitioners' use of gained insights and potential impact on the infrastructure.	How could the gained insights from the workshop help with your currently faced challenges in transforming your infrastructure?
			How did you further proceed with the gained insights from the workshop?

Figure 6.5: Data collection questionnaire for supportability-related evaluation criteria

Adoption concerning Evaluation Criteria

The third cluster of evaluation criteria concerns adoption-related criteria. We adapt these evaluation criteria to the evaluation context by utilizing the practitioners' recommendation willingness as a proxy for measuring the GaaPIAM's adoption likelihood by the practitioners.

In the second step of the evaluation concept configuration, we choose an adequate evalua-

tion methodology. Since capturing GaaPIAM's actual adoption by the practitioners exceeds the temporal scope of our evaluation process, we research for an adequate evaluation methodology that can capture the recommendation willingness and, in consequence, allow us to infer a high adoption likelihood in the long term. The behavioral science literature hosts such a methodology fulfilling our evaluation requirements for adoption-related evaluation criteria. Reichheld's Net Promoter Score is a methodology that primarily aims to measure customer satisfaction and, hence, loyalty expressed through returning buying or engagement behavior [48, 49]. Reichheld additionally does not introduce any restrictions on what subject the Net Promoter Score can be applied to but rather proposes a wide range [49]. Thus, we utilize it to estimate if practitioners would re-use the GaaPIAM in the long term. Furthermore, the Net Promoter Score calculation is based on the recommendation willingness of the subjects of interest, in our case, the practitioners [48, 49]. This aligns ideally with our context challenge solution to utilize the practitioners' recommendation willingness.

Regarding the data collection configuration, we follow the scientific guidelines for formulating Net Promoter Score-based questions [48, 49]. Consequently, when questioning the practitioners regarding their recommendation willingness, we use the word "recommend" in the question formulation. Accordingly, we created three Net Promoter Score questions derived from each adoption-related evaluation criterion. Furthermore, we also include a Net Promoter Score question regarding the general recommendation willingness of GaaPIAM's instantiation, hence, the workshop.

Configuring the evaluation setting, we pose each adoption-related and the general Net Promoter Score question directly after the workshop. We use a digital tool to pose the Net Promoter Score question anonymously. Figure 6.6 displays all formulated Net Promoter Score questions and their temporal settings.

Evaluation Criteria	Evaluation Timing	Data Collection Goal	Question
Non-Specific	Directly after the instantiation	Practitioners' overall willingness to recommend GaaPIAM's instantiation	How probable is it that you recommend the workshop to a colleague?
Adoption of GaaPIAM due to increased understandability of GaaP	Directly after the instantiation	Practitioners' willingness to recommend GaaPIAM due to increased understandability of GaaP	How probable is it that you recommend the workshop due to the increased understandability of GaaP to a colleague?
Adoption of GaaPIAM due to the identification of gaps in infrastructure's architecture	Directly after the instantiation	Practitioners' willingness to recommend GaaPIAM due to identification of gaps in infrastructure's architecture	How probable is it that you recommend the workshop due to the identification of gaps in infrastructure's architecture to a colleague?
Adoption of GaaPIAM due to the development of actionable proposals for improving infrastructure	Directly after the instantiation	Practitioners' willingness to recommend GaaPIAM due to the development of actionable proposals for improving infrastructure	How probable is it that you recommend the workshop due to the development of actionable proposals for improving infrastructure to a colleague?

Figure 6.6: Data collection questionnaire for adoption-related evaluation criteria

7 Evaluation Results

In the following section, we present the results from applying the evaluation concept to the instantiation of the GaaPIAM. The instantiation of the GaaPIAM occurred within two workshops with two different German federal agencies. Germany has mandated by law the digitalization of public services called "Onlinezugangsgesetz," abbreviated "OZG." [29, 30]. Hence, Germany is in a transitioning and transforming phase regarding its digital public service offerings and respective digital infrastructures. However, this digitalization process of public services requires the development of new systems and infrastructures. The practitioners who participated in our workshops are part of the institutions which are tasked with the endeavor to develop these new systems.

The below Figure 7.1 displays an overview of the evaluation results.

Evaluation Criteria	Evaluation Results	Data
Degree of effectiveness in increasing the understandability of GaaP	Strong support	Survey data and practitioners' statements show a strong indication that GaaPIAM increases the understandability of GaaP
Degree of effectiveness in identifying gaps in the infrastructure	Partial support	Survey data shows partial support. Following practitioners' statements, they identified missing roles, governance drawbacks, and new platform-related architectural concepts for their infrastructure
Degree of effectiveness in developing actionable proposals for improving infrastructure	Support	Survey data shows support. Following practitioners' statements, they developed a better-perceived roles-structure for their infrastructure, architecture complexity reduction potentials through outsourcing of components into the ecosystem, and conceptual vocabulary for stakeholder communication purposes of GaaP
Supportability (extent of support) related Evaluation Criteria	Support	The practitioners' statements show support. Practitioners utilized generated insights to draw a clearer roles-structure for their infrastructure, incorporated openness and co-participation in their architectural governance, existing platform-oriented ideas were endorsed, and the acquired conceptual vocabulary was used in the formulation of requirements
Adoption of GaaPIAM due to increased understandability of GaaP	Low support	Net Promoter Score: 33
Adoption of GaaPAM due to identification of gaps in infrastructure's architecture	No support	Net Promoter Score: 0
Adoption of GaaPIAM due to the development of actionable proposals for improving infrastructure	No support	Net Promoter Score: 0

Figure 7.1: Overview of evaluation results

General workshop design

The workshops took place in February 2023. The first workshop was conducted online

using Microsoft Teams, and the second workshop was conducted in person at the offices of the government organization. To ensure anonymity and non-traceability, we do not disclose the organizations' locations. All participants consented to record the workshop for research and evaluation purposes. The participants provided beforehand a blueprint visualization of their respective digital infrastructure which then was the central subject for applying the GaaPIAM. We utilized Miro as a collaboration, documentation, and artifact instantiation tool. Miro is a digital whiteboard cloud application. Figure 7.2 displays the template used for the workshops without the practitioners' infrastructure. The actual filled Miro board was provided to the participants after the workshop for documentation purposes.

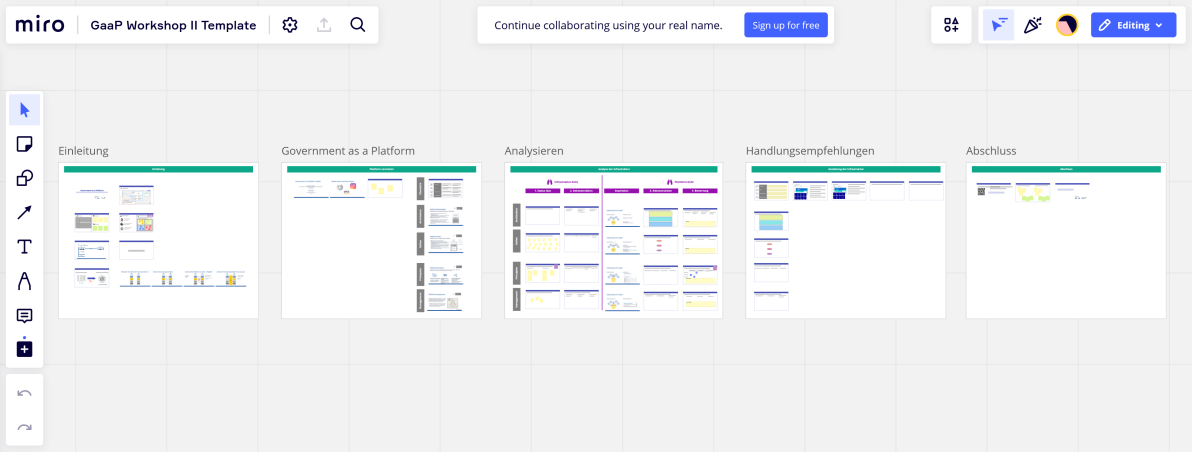

Figure 7.2: Workshop template in Miro

The workshop is structured along five phases, which can also be seen on the Miro Board 7.2. In the first phase, the introduction, we introduced ourselves and the participants and provided a general overview of the workshop and topic. The second phase was designed to deep-dive into the GaaP approach organized along the four dimensions of the GaaPIAM with the goal of increasing the understandability of the GaaP approach. The third phase concerned the analysis of the participant's infrastructure by applying the GaaPIAM. Subsequently, in the fourth phase, we infer actionable proposals for improving the practitioners' infrastructure based on the previous analysis. We then concluded the workshop with quick general feedback from the participants and urged them to fill out our digital evaluation survey. The workshop was held in an interactive and participative manner, where the researchers, as well as the practitioners, collaboratively engaged while applying the GaaPIAM.

Figure 7.3 displays the five phases through the workshop's general agenda.

Phase	Duration
Welcoming & Introduction	**15 min**
Part 1: Government as a Platform	**30 min**
Part 2: Analysis of Infrastructure	**120 min**
Part 3: Development of Actionables	**30 min**
Survey & Feedback	**15 min**

Figure 7.3: General workshop agenda

Data collection tools

Besides the recordings and transcriptions of the workshop, we further collected field notes during the workshop. We performed the semi-structured pre- and post-interviews verbally with the participants as described in the developed evaluation concept in chapter 6. The data collection methodologies and procedures are further thoroughly described in chapter 4. However, regarding the survey, which was conducted directly after the workshop, we utilized Microsoft Forms, which is not discussed in the previous chapters. Microsoft Forms provides a complete user interface paired with a pre-installed builder for the Likert scale and Net Promoter Score questions, plus a built-in analysis dashboard. Again we used the developed evaluation concept from chapter 6 to configure the questionnaire using these Microsoft Forms' features. More precisely, all Likert scale questions from Figure 6.4 and Net Promoter Score questions meant for directly after the workshop from Figure 6.6 were translated into German and utilized in the survey. The participants could access the survey via a link or QR-Code in the last phase of the workshop.

Figure 7.4 displays an excerpt from the Microsoft Forms survey.

Figure 7.4: Excerpt from Microsoft Forms survey for workshop participants

Labels and acronyms in this chapter

To describe from where the quotes and data originate in the following sections, we labeled each workshop, practitioner, and practitioners' infrastructure accordingly using anonymized acronyms. Table 7.1 describes the acronyms and labels and their respective meaning.

Acronyms and Labels	Meaning
WS-1	The first Workshop
WS-2	The second workshop
PRE-I-1	The first pre-interview following WS-1
PRE-I-2	The second pre-interview following WS-2
POST-I-1	The first post-interview following WS-1
POST-I-2	The second post-interview following WS-2
P1-WS-1	The first practitioner from the first workshop
P2-WS-1	The second practitioner from the first workshop
P1-WS-2	The first practitioner from the second workshop
X-WS-1	The core infrastructure of interest from the first workshop
Y-WS-1	A secondary infrastructure of interest from the first workshop
Z-WS-2	The core infrastructure of interest from the second workshop

Table 7.1: Labels and acronyms legend

7.1 Corroboration of Evaluation Assumptions

In the previous chapter 6, we identified the practitioners as proxies for the evaluation due to the inaccessibility of the governments' actual digital infrastructures. Consequently, the

underlying general assumption is that the practitioners who participate in our workshop actually have the ability and authority to impact the infrastructure's architecture and design. Therefore, before performing the evaluation, the evaluation concept foresees corroborating this assumption to ensure rigor and relevance [47]. Hence, we utilized the questionnaire 6.3 from the evaluation concept, which we asked during the semi-structured pre-interview.

The first practitioner P1-WS-1 from the first workshop was a "software architect" [PRE-I-1, 00:05:22] responsible for the infrastructure "X-WS-1" (referred to as X-WS-1 to maintain anonymity) [PRE-I-1, 00:05:22] and bringing "all interests somehow together" [PRE-I-1, 00:05:22]. The second practitioner, P2-WS-1 from the first workshop, was "Head of Department," [PRE-I-1, 00:08:39] and his tasks were comprised of the management of the "X-WS-1-landscape, development of the X-WS-1-Landscape [...] and integration of X-WS-1 in the area of architecture and development of Y-WS-1" [PRE-I-1, 00:08:39], where Y-WS-1 is the pseudonym for another infrastructure.

The practitioner P1-WS-2 from workshop 2 was a "Product Manager" [PRE-I-2, 00:00:45] responsible for his infrastructure, which we refer to as Z-WS-2. Although from an architectural-decision perspective, the practitioner stated that he conducts "architecture on a more abstract level" [PRE-I-2, 00:00:45], he has a "veto right" [PRE-I-2, 00:00:45] on all architectural decisions.

Hence, we conclude that the evaluation concept's general assumption to utilize the practitioners as a proxy can be considered corroborated regarding all practitioners over all workshops because they have a substantial influence on the infrastructure's architecture.

Moreover, regarding the participants' motivation, the practitioners from workshop WS-1 stated that they "want to learn something new and exchange ideas" [PRE-I-1, 00:17:33] and obtain a "new perspective different from [their] current one" [PRE-I-1, 00:18:37]. Additionally, participant P1-WS-1 stated that he wishes for a tool that supports "more this de-composition, to decompose problems into sub-problems. [...] There is, of course, the modularization approach and defining interfaces, but something is missing" [PRE-I-1, 00:28:37].

Practitioner P1-WS-2 furthermore described his motivation to "further design the product" [PRE-I-2, 00:12:32], especially regarding "how the Z-WS-2 can be seen as a platform and how it integrates itself in a bigger platform ecosystem" [PRE-I-2, 12:32]. Regarding the necessity of additional tool support, the practitioner said that "looking at other countries" [PRE-I-2, 00:10:11] and making their experiences and learnings "compatible with Germany and its federalism" [PRE-I-2, 00:10:11] can be beneficial.

From these practitioners' statements, we can conclude that the GaaPIAM might be relevant to them as, on the one hand, the method helps to decompose the infrastructure into modular components [18] and, on the other hand, builds upon literature which describes how countries successfully approach GaaP [6, 8, 7, 18].

7.2 Effectiveness of GaaPIAM

Following the evaluation concept, evaluating the effectiveness of the GaaPIAM is coupled with accepting or rejecting the three utility hypotheses defined in chapter 6. As determined in the evaluation concept, to assess the utility hypotheses, we analyze the collected data from

the survey and the open questions from the post-interview.

Evaluation of Utility Hypothesis 1: The GaaPIAM will better increase the understandability of the GaaP approach to the practitioners than without it

Regarding utility hypothesis 1, the survey data over all workshops consistently show strong support as all practitioners strongly agree with the respective Likert scale question. Additionally, in workshop 1, without our request, the practitioners strongly engaged, gave positive feedback, and transferred the contents to their own infrastructure during the explanation of the GaaP approach using the GaaPIAM's dimensions. For example, practitioner P1-WS-1 stated regarding the infrastructure's architecture that he is "speculating the whole time how strongly the core is built in [their] infrastructure and where the boundary resources and ecosystem would be" [WS-1, 01:09:28]. Subsequently, also about the infrastructure's architecture, he said that it "helps to put it into another frame and think differently about it" [WS-1, 01:09:54]. Furthermore, the same practitioners stated that the understandability of platform roles is "totally logical and comprehensible" [WS-1, 01:15:38]. Considering the understandability of the platform principles, he affirmed that it is "also understandable" [WS-1, 01:20:02]. A similar engagement and feedback could be observed in workshop WS-2. There, the practitioner P1-WS-2 stated regarding the architectural dimension of GaaP that he could "comprehend [it]. There are things you must look at, things you partially must think through" [WS-2, 00:20:02]. Additionally, he pointed out that he "liked well the derivation [...] of Co-Creation" [WS-2, 00:29:13]. The same practitioner also showed strong engagement by transferring the acquired knowledge about the GaaP approach to a real-life example concerning Atlassian's business model and components. In this context, he noted, "wonderful example with Atlassian: Jira, Jira Plug-In, Jira Users" [WS-2, 00:23:08].

In the subsequent post-interviews, the practitioners from all workshops affirmed that the workshop helped them to understand the GaaP approach better. For example, practitioner P1-WS-1 stated that he "understood it better" [POST-I-1, 00:00:28] while his colleague, practitioner P2-WS-1, also stated that it helped him to gain a "better understanding of GaaP" [POST-I-1, 00:01:22]. Additionally, the practitioner P1-WS-2 from the second workshop stated that the workshop helped him acquire "conceptual vocabulary and a mental model" [POST-I-2, 00:00:21].

Hence, we conclude the first utility hypothesis can be accepted based on the participants' evaluation and statements. Thus, the GaaPIAM increases the understanding of the GaaP approach.

Evaluation of Utility Hypothesis 2: The GaaPIAM will better enable practitioners to identify gaps in their infrastructure's architecture than without it

Regarding utility hypothesis 2, the survey data over all workshops shows mixed results. One practitioner strongly agrees with the hypothesis, while another practitioner agrees and another partly agrees. Due to the anonymity of the survey, we cannot trace back which

practitioner from which workshop agreed more or less. However, during the workshops, the application of the GaaPIAM could enable practitioners to reveal new insights regarding gaps and optimization potentials in their infrastructure. Practitioner P2-WS-1, for example, stated that they "so far did not look at it at all. [...] Now we come to the topic of standards. It still must be defined in what manner the registry answers so that the requesting entity, the data consumer, can also work with the answer. That means there must be beyond this [...] more standards" [WS-1, 02:15:07]. Additionally, P2-WS-1 stated that "there exist potentially more core elements, that [he] would like to include. [...] These are currently not in the context of [their] work. But there are some central services that [he] would like additionally to offer to make it for all involved parties easier to build an ecosystem" [WS-1, 02:18:36]. Furthermore, the same practitioner stated in another context that "in a perfect world [he] would offer [an infrastructure component] as a module so that the portal can be translated into a unified UI" [WS-1, 03:37:56]. The other practitioner P1-WS-1 from the same workshop also stated that "this participation and co-creation is a desirable thing, where [they] also want to get there. However, there is still much to do" [WS-1, 03:05:45]. In the second workshop, the practitioner could quantitatively identify fewer gaps in his infrastructure. However, he noted that regarding the "degrees of freedom and decision-making power" [WS-2, 01:18:34] they can "not dictate because they are not [their] own master over [their] platform. That is maybe an important insight" [WS-2, 01:18:34].

In the subsequent post-interviews, the initial observations from the workshops and surveys are predominately reflected also in the post-interviews. Practitioners from the first workshop incorporated the derived insights to obtain a clearer role structure and decrease complexity by, for example, "categorize[ing] software components. [...] as it helps [the practitioner] to say that this a component, where [they] do not need to make provisions because those are things people can later romp about and contribute themselves" [POST-I-1, 00:13:12]. Furthermore, practitioner P2-WS-1 added that a "user distinction: Who is the owner? Who is a contributor? That is something [they] momentarily actively work on, to draw a clear picture because [they] took the complete responsibility for the "X-WS-1" as a whole" [POST-I-1, 00:17:24]. However, as already observed during the second workshop, practitioner P1-WS-2 incorporated fewer improvements in his infrastructure as he quantitatively generated fewer improvements during the workshop. Nevertheless, he stated that it reinforced certain ideas, such as "certain decouplings, that [they] already planned with the API gateway as a component" [POST-I-2, 00:10:39].

Hence, we conclude the second utility hypothesis can be partially accepted based on the participants' evaluation and statements, particularly from the first workshop. Thus, the GaaPIAM can enable practitioners to identify gaps in their infrastructure's architecture to a certain degree.

Evaluation of Utility Hypothesis 3: The GaaPIAM will better enable practitioners to develop actionable proposals for improving their infrastructure than without it

Regarding utility hypothesis 3, the survey data over all workshops show support as all

practitioners agree with the respective Likert scale question. This can also be corroborated by the actionable proposals for infrastructure improvements the practitioners generated in the two workshops. Particularly in the first workshop, practitioners applying the GaaPIAM to their infrastructure could generate multiple actionable proposals and design ideas regarding the roles and responsibilities revolving around their infrastructure. For example, during the task of identifying all relevant actors in their infrastructure and then recomposing them to typical platform roles, practitioner P2-WS-1 stated that "this differentiation [between complementors and owner] [he] likes" [WS-1, 02:27:30], particularly, re-assigning certain "[actors] rather into the complementors list" [WS-1, 02:27:30]. After completing this task, the practitioner P1-WS-1 stated that "this looks more beautiful than it actually is" [WS-1, 02:30:15], and the practitioner P2-WS-1 subsequently added that "if [they] implement that [roles] illustration of how [they] generated it, it would help [them]" [WS-1, 02:31:41]. Furthermore, the latter practitioner also stated that he "is always trying to categorize those states, and [he] likes them as complementors quite good" [WS-1, 02:35:51]. Practitioner P1-WS-1 corroborated that statement by adding that "this complementor role is maybe for the states exactly what they actually need and want" [WS-1, 02:32:30]. Finally, concluding the actionable proposals generated concerning roles, the practitioners strongly affirmed the researcher's suggestion to "clearly and unambiguously define the platform owner" [WS-1, 03:47:55] as well as "equip him with a strong mandate" [WS-1, 03:50:21]. Moreover, concerning the infrastructure analysis regarding the adherence to platform principles, practitioner P2-WS-1 stated that he "finds this idea great [...] not to build their own portals. [They] could let others build them" [WS-1, 03:29:50]. Another re-design idea generated by the practitioner P2-WS-1 through the application of the GaaPIAM is to locate "[an infrastructure component] into the ecosystem. [He] would offer the data openly [...] and let whoever wants to implement it" [WS-1, 03:31:02]. Likewise, in the second workshop, practitioner P1-WS-2 could identify actionable proposals to improve his infrastructure. For example, he stated that a certain component of his infrastructure Z-WS-2 should not have authentification to retrieve data, but rather this data could be provisioned as "open data" [WS-2, 00:46:02] as they "actually do not require authentification for this service" [WS-2, 00:46:02]. Another insight, which subsequently generated an actionable proposal, concerns the roles within the practitioner's infrastructure. According to the practitioner P1-WS-2, certain stakeholders "understand themselves as [their] bosses, although they are not" [WS-2, 01:46:32], resulting in the inability to effectively "set provisions, as [they] lack the leverage" [WS-2, 01:55:29]. Hence, as an actionable, the practitioner P1-WS-2 recognizes that it is "a question of mandate and ownership" [WS-2, 01:56:46] regarding his infrastructure.

Furthermore, the results from the subsequent post-interviews demonstrate how the practitioners from all workshops already incorporated insights gained through the application of the GaaPIAM in their work. For example, practitioners from the first workshop incorporated insights from the workshop into their "architectural governance [...], where [they] said these are the guidelines with which the system must be developed including the core-system as well as the other systems" [POST-I-1, 00:18:47]. Particularly the idea that if complementors "think [they] need something, [they can] build the system [and] offer it to others so that they can use it" [POST-I-1, 00:18:47] was encompassed in the new architectural governance. In the

same context regarding the latter, practitioner P1-WS-1 states that "in the past [he] would have said [they] have to determine the provisions. And now, when [he] say[s] this is freely designable, [he] perceive[s] it as a significant added-value and then [he] know[s] he can give development wiggle-room for others" [POST-I-1, 00:16:53]. Moreover, practitioner P2-WS-1 stated in the context of obtaining a clearer role structure within their infrastructure that they have "established a responsible for all core components [of the infrastructure]" [POST-I-1, 00:18:47]. On the other hand, the practitioner P1-WS-2 from the second workshop stated that he obtained through the workshop "conceptual vocabulary and a mental model" [POST-I-2, 00:00:21], which helped him to depict a more "stringent requirements management [...] which describes also the requirements well" [POST-I-2, 00:12:58].

Hence, we conclude the third utility hypothesis can be accepted based on the participants' survey data and statements. Thus, the GaaPIAM can better enable practitioners to develop actionable proposals for improving their infrastructure.

7.3 Supportability of GaaPIAM

The second pillar of the evaluation concept concerns the ability of the GaaPIAM to support the practitioners to impact the infrastructure. For that purpose, we followed our developed evaluation concept from chapter 6 and conducted a pre-interview as well as post-interview.

Identified challenges in pre-interviews

In the pre-interview, we identified current challenges the practitioners face in transforming their infrastructures. Regarding the practitioners from the first workshop WS-1, two main challenges were stated. The first challenge concerns the complexity of the system in terms of its architecture due to high cohesion, regulatory requirements, and organizational issues. For example, the practitioner P2-WS-1 stated that "[their] infrastructure how [they] momentarily plan feels ten times more costly and ten times so complex as it should be" [PRE-I-1, 00:21:10]. The practitioner further affirmed his colleague's statements by adding that their infrastructure "is a highly complex technical system that is more intermeshed than it should be. But not only from a technical perspective but also from a judicial, organizational, and political one" [PRE-I-1, 00:28:25]. The second challenge the practitioners face regards the organization, especially the role distinction within the infrastructure. Regarding the latter challenge, practitioner P2-WS-1 stated that "it is a big system [...] and it is absolutely open who can build something and who can operate it. Because everybody wants to play a role, but no one wants to take responsibility. And exactly that is the problem we have in Germany: This clear distinction between project setup of a system and operations of a system" [PRE-I-1, 00:21:10].

On the other hand, the practitioner P1-WS-2 faced similar challenges regarding the roles within their infrastructure. For example, he lamented the lack of a strong mandate for himself by stating that they "do not have a position where they can easily mandate. [They] still have to 'harmonize' within their [organizational] environment, which is a costly and tedious

process" [PRE-I-2, 00:14:56]. Additionally, he stated that "the governance structures that exist momentarily for the development of the product are not quite suitable as they need to do things also quite exploratively" [PRE-I-2, 00:13:20].

Workshop insights supporting practitioners in facing their challenges identified in post-interviews

After the workshop was conducted and the practitioners were exposed to the GaaPIAM, we again interviewed them over a temporal distance of around three to four weeks from the workshops. The focus of these post-interviews lay on the exploration of how the practitioners coped with their initial challenges after generating insights and improvements regarding their infrastructures by applying the GaaPIAM during the workshops.

Regarding the first workshop, evidence can be derived from the post-interview displaying how the workshops helped support the practitioners to cope with their initial challenges and hence impact their infrastructure. For example, practitioner P1-WS-1 stated that the workshop helped him to "categorize software components. [...] as it helps [him] to say that this a component, where [they] do not need to make provisions because those are things people can later romp about and contribute themselves" [POST-I-1, 00:13:12]. He later added that to the latter, that "this is very valuable, because in the past [he] would have said [they] have to determine the provisions. And now, when [he] say[s] this is freely designable, [he] perceive[s] it as a significant added-value and then [he] know[s] he can give development wiggle-room for others" [POST-I-1, 00:16:53]. In this case, he clearly states that through the insights of the application of the GaaPIAM, he now approaches more openly the determination of responsibilities for the development of components by allowing third parties to co-participate and contribute more. This helps the practitioner to better cope with the challenges captured in the pre-interview, as they state that they now focus more on core components and let others develop certain components, thus, reducing their overhead and complexity. Additionally, it helps them cope with the unclear role structure regarding who develops and operates what by letting other parties develop and provide additional components. Consequently, they are shifting development and operations responsibilities regarding the latter to third parties. Furthermore, when we asked the practitioners if there were concrete insights derived from the workshop affecting them or their infrastructure, they stated that they established an "architectural governance [...], where [they] said these are the guidelines with which the system must be developed including the core-system as well as the other systems" [POST-I-1, 00:18:47]. Additionally, this architecture governance included a concrete insight derived from the workshop as the practitioner P2-WS-1 stated that when third parties "think [they] need something, [they can] build the system [and] offer it to others so that they can use it. That way [he has], for example, again two competing systems that do the same from two different entities" [POST-I-1, 00:18:47]. This corroborates the above statements that the application of the GaaPIAM supports them with the challenges they face by enabling the practitioners to let other entities develop software components interoperably with their core infrastructure. By doing so, the practitioners reduce the complexity of their

system as they outsource complementary components from their core and simultaneously shift responsibilities to external entities. Thus, the role structure within the system becomes clearer too. Moreover, regarding the latter, the practitioner P2-WS-1 added that they are utilizing the role structures from the workshop to conduct a "user distinction: Who is the owner? Who is a contributor? That is something [they] momentarily actively work on, to draw a clear picture because [they] took the complete responsibility for the "X-WS-1" as a whole" [POST-I-1, 00:17:24]. Particularly, they have "established a responsible for all core components [of the infrastructure]" [POST-I-1, 00:18:47].

In the second workshop WS-2, the practitioner P1-WS-2 stated that the workshop primarily "reinforced ideas that [they] already had" [POST-I-2, 00:10:39]. Particularly considering his challenge concerning the role structure, he stated using an example of a newly introduced "user group that it will play a role in the whole as a communication plenum and as an exchange and source of requirements. That [...] was strengthened through [the workshop]" [POST-I-2, 00:10:39]. Furthermore, upon the question of how the model of the complementor role identified during the workshop impacted his work so far, he added that "in the context of [their] internal thinking, it already happened, that [external entities] acquired this role" [POST-I-2, 00:16:05]. However, a major aspect identified during the post-interview with practitioner P1-WS-2 that was not mentioned in the pre-interview beforehand concerns the acquisition of "conceptual vocabulary and a mental model" [POST-I-2, 00:00:21] through the application of the GaaPIAM in the workshop. Consequently, the workshop supported the practitioner P1-WS-2 to utilize this "consistent conceptual vocabulary throughout the requirements," [POST-I-2, 00:14:46] enabling introducing a more "stringent requirements management [...] which describes also the requirements well" [POST-I-2, 00:12:58]. Yet about the same context, the practitioner P1-WS-1 from the first workshop WS-1 stated that "[he] know[s] now that there is this conceptual vocabulary but [he] do[es] not have it in [his] vocabulary. It also does not help because when [he] communicates it to somebody, then the other one has to be enlightened first. Him [he] must first explain everything [...]" [POST-I-1, 00:06:14]. Hence, the acquisition of conceptual vocabulary is not an important generated insight for the practitioners from the first workshop WS-1.

Additionally, practitioners from both workshops lamented that the time frame between the workshop and post-interview is too small "to go into details" [POST-I-2, 00:06:34]. Furthermore, especially practitioners from the first workshop expressed their desire for more material like a "checklist [...] to repeatedly reflect upon" [POST-I-2, 00:10:05].

Although considering the way how the application of the GaaPIAM supported the practitioners differently and the limited time frame between workshops and post-interviews, following the above statements about how the application supported the practitioners in their challenges, it can be concluded that the Method supports the practitioners in facing them.

7.4 Adoption of GaaPIAM by practitioners

Lastly, an integral part of the evaluation concept from chapter 6 is the evaluation of GaaPIAM's adoption by practitioners. The evaluation concept foresees capturing the adoption by using

the recommendation willingness of the practitioners by utilizing the Net Promoter Score directly after the workshop in a survey format.

Before evaluating the adoption by the practitioners for each evaluation criterion as intended in the evaluation concept, we additionally asked the practitioners to state their overall recommendation willingness regarding the workshops to obtain a more general insight regarding the adoption willingness. Here, one practitioner scored the workshop with a seven, and another gave the workshop an eight. Both scores indicate that these two practitioners have a more neutral stance towards the workshop and would most probably not retain the workshop nor actively recommend it [48, 49]. Nevertheless, one practitioner scored the workshop with a value of nine, which translates that he would actively recommend the workshop [48, 49]. Consequently, the overall Net Promoter Score results to 33 as one out of three would actively promote the workshop.

Hence, only a partial to low adoption willingness can be inferred regarding the overall adoption of the GaaPIAM.

Evaluation of Adoption of the GaaPIAM by practitioners due to the increased understandability of GaaP

The scores regarding the adoption of the GaaPIAM by practitioners due to increased understandability of GaaP display a similar result as the overall recommendation willingness of the workshop. Hence, two practitioners scored the respective Net Promoter Score question with an eight, meaning they are passives [48, 49]. However, one of the practitioners scored the same question with a nine and, thus, can be considered a promoter [48, 49]. Consequently, the overall Net Promoter Score results to 33 as one out of three would actively promote the workshop due to increased understandability of GaaP.

Hence, only a partial to low adoption willingness can be inferred regarding adopting the GaaPIAM due to the increased understandability of GaaP.

Evaluation of Adoption of GaaPIAM by practitioners due to the identification of gaps in infrastructure's architecture

The scores regarding the adoption of the GaaPIAM by practitioners due to the identification of gaps in the infrastructure's architecture are homogeneous, as two practitioners gave a score of eight and another one a score of nine. Hence, all practitioners can be considered passives [48, 49]. Consequently, the Net Promoter Score amounts to zero. However, practitioner P1-WS-1 from workshop 1 stated that it is a "helpful point, when [...] you repeatedly question the structure [...] and from time to time also bring those test questions" [WS-1, 02:12:33] regarding the decomposition and recomposition of infrastructure components. This statement can indicate a willingness to re-use the GaaPIAM over time and, thus, simultaneously indicate its adoption.

Nevertheless, following the respective Net Promoter Scores and the only one identified statement indicating adoption willingness, a very weak to no adoption willingness of the

GaaPIAM due to the identification of gaps in the infrastructure's architecture can be concluded.

Evaluation of Adoption of the GaaPIAM by practitioners due to the development of actionable proposals for improving infrastructure

Similarly to the above evaluation, the adoption of the GaaPIAM by practitioners due to the development of actionable proposals for improving their infrastructure, there are no promoters nor detractors. Hence, two practitioners chose a Net Promoter Score of seven, while another practitioner a score of eight. Therefore, all practitioners can be categorized as passives with an overall Net Promoter Score of zero [48, 49]. Additionally, during the workshops, no clear and distinct indication could be distilled that could indicate a long-term adoption willingness.

Hence, following the respective Net Promoter Scores, no adoption willingness of the GaaPIAM due to the development of actionable proposals for improving the infrastructure can be concluded.

8 Discussion

The purpose of the thesis at hand was to develop a suitable evaluation concept for rigorously evaluating the GaaPIAM by Kuhn et. al [18]. Within this scope, we identified several key findings and limitations regarding the GaaPIAM and the respective developed evaluation concept, which we both critically assess in the following.

8.1 Key Findings and Limitations regarding GaaPIAM

The evaluation of the GaaPIAM yielded in total concise results showing that the GaaPIAM supports practitioners to understand better the GaaP approach, identify gaps and optimization potentials in their infrastructure, and generate concrete actionable proposals for improvement of their infrastructures.

In the following, we reflect upon and assess the key findings and limitations regarding the GaaPIAM.

The GaaPIAM supports practitioners differently approaching GaaP and generates different added values for them

The evaluation shows that the kind of insights generated differs between practitioners. Consequently, also the added value from practitioners applying the GaaPIAM differs. For example, the practitioners from the first workshop benefited from the generated insights regarding the use of open and co-participative architectures in their infrastructure. Accordingly, they incorporated these architectures and underlying principles into their architectural governance to reduce the complexity of the core system by allowing the development of complementary non-core components without posing provisions. However, the primary insight generated by the practitioner from the second workshop revolves around the acquired conceptual vocabulary supporting him to define more coherent and unified requirements and communicate the platformization idea better to stakeholders.

Indication for such diversion regarding the differently generated insights might originate from the different individual and organizational backgrounds. Although all practitioners are part of German federal institutions tasked with the digitalization of public services, a clear distinction in their prior knowledge concerning platforms as well as organizational culture can be derived. For example, the practitioners from the first workshop are affiliated with a more traditional and older federal institution, and both stated in the pre-interviews that they have less platform knowledge and "no idea of [GaaP]" [PRE-I-1, 00:15:05] & PRE-I-1, 00:15:13]. Hence, it might insinuate why these practitioners generated quantitatively more insights

and, additionally, primarily insights regarding GaaP principles and their implication on their architecture. On the other hand, although the practitioner from the second workshop stated that he has "limited" [PRE-I-2, 00:02:35] knowledge regarding the GaaP approach, he rates his general platform knowledge as "eight" [PRE-I-2, 00:20:50] on a scale of ten. Additionally, he said during the workshop that his institution is "in terms of openness [...] very value-driven, and things like transparency and whatnot lay well on top of each other" [WS-2, 02:32:35]. Thus, certain platform principles are already part of the institution's culture. That might provide a first insight into why this practitioner felt that the workshop mostly "reinforced ideas that [they] already had" [POST-I-2, 00:10:39] and primarily provided him with tools to communicate and formulate better requirements based on the acquired conceptual vocabulary.

Hence, the GaaPIAM supports practitioners differently and generates different added values for them, possibly based on the organizational values' similarity to platform principles as well as practitioners' prior knowledge regarding platforms in general.

The GaaPIAM supports practitioners to analyze and improve the roles-structure and respective relations within their infrastructure

Despite the GaaPIAM supporting practitioners differently and generating different added values for them, the Method depicts a common deficit in the practitioners' infrastructure regarding the roles-structure and their respective relations. The practitioners from all workshops lamented the lack of a strong mandate for the infrastructure owners as well as the lack of a unified categorization for contributing third parties to the infrastructure. Particularly regarding the latter, it often led to tensions between institutional entities regarding requirements and decision rights, according to the practitioners. Furthermore, often no clear responsibilities regarding core functions were assigned, such as the development and the operations of the infrastructure's core components according to the practitioners' statements from the first workshop. However, the role structure introduced by the GaaPIAM and the insights generated by its subsequent application revealed that the role of the complementor is a suitable construct to categorize the contributing parties and better define their responsibilities.

Hence, the GaaPIAM supports practitioners in analyzing and improving the structure and relations of the roles within their infrastructure.

The GaaPIAM supports uncovering legislative and regulatory barriers affecting the implementation of the GaaP approach

Another rather unexpected insight generated by the GaaPIAM is its potential to demonstrate implementation barriers of the GaaP approach due to regulatory requirements affecting technical and organizational aspects of the practitioners' infrastructure. These implementation barriers are especially observable in the first workshop. One practitioner even states that the "existing law shoots [them] in the knee" [PRE-I-1, 00:21:10]. According to this practitioner, this is due to the fact that the "already existing legislation [...] is written in such a way that it makes it impossible to implement things like an information technology professional would

build them" [PRE-I-1, 00:21:10]. Thus, resulting in an infrastructure "ten times more costly and ten times so complex as it should be because [they] just try to follow some legislations" [PRE-I-1, 00:21:10]. In a more concrete example, one practitioner formulated the wish based on a generated insight to outsource "the data protection cockpit in the ecosystem. [He] would provide the data openly [...] and whoever wants can implement it" [WS-1, 03:31:02]. However, later he admitted that there are "legal issues to resolve" [WS-1, 03:43:32] in order to implement the data protection cockpit in that way.

Hence, the GaaPIAM supports the practitioners in uncovering legislative and regulatory barriers that must be eliminated to not affect the implementation of the GaaP approach.

The GaaPIAM requires further instantiation mediums to better support practitioners

The GaaPIAM's instantiation is primarily designed as a one-time event in form of a workshop. While this provides the practitioners with a good impulse and already generates valuable insights, it further requires ways to refresh and communicate the knowledge and also repeatedly compare core insights generated by the GaaPIAM with the practitioner's infrastructure cost-efficiently. For example, one of the practitioners from the first workshop stated that he would need "to perform the workshop again or communicate it in a way so that everybody understands what it is about" [POST-I-1, 00:10:05]. The other practitioner added in the same context that "this first workshop was the management pitch to plant the thought and make the decision makers want it. Then the next step would be to involve the working population" [POST-I-1, 00:27:30]. Additionally, another practitioner added that he "needs a refresh which will not take place in form of a workshop where they take a lot of time for it, but instead the refresh can only happen by having a written reference" [POST-I-1, 00:13:12]. Moreover, both practitioners mentioned the need for some form of "checklist [...] where [they] can reflect upon" [POST-I-1, 00:10:05] or a "review [...] where [they] look at what was achieved, what failed, what needs adjustment [...] to build a platform. [...] leading to a continuous accompanying process" [POST-I-1, 00:27:30]. Likewise, the practitioner from the second workshop corroborated the idea of having a reference "to form the architecture of the product in order to speak with the colleagues about the same thing and not leave too much room for interpretation" [POST-I-2, 00:27:29]. He also added that there is a "socio-technical aspect [regarding] how to sell [the platform idea] to the stakeholders, how to address them correctly, how to find out what is important to them and how to address their concerns" [POST-I-1, 00:28:46] which he also would like better guidance on.

Hence, the GaaPIAM is an adequate tool to demonstrate to the practitioners what is possible by platformizing their infrastructure and generating first insights on how to achieve this. Nevertheless, it requires further ways to instantiate the Method, helping refresh and communicate the GaaP approach as well as cost-efficiently review the current platformization progress of the practitioners' infrastructures against a reference architecture.

8.2 Key Findings and Limitations regarding GaaPIAM's evaluation concept

The evaluation concept in this thesis was developed to fit the evaluation of the GaaPIAM by utilizing primarily literature from the Design Science Research Evaluation theory. The application of this evaluation concept yielded the above results in chapter 7.

Therefore, in the following, we reflect upon and assess the key findings and limitations regarding the evaluation concept to better understand and interpret the yielded evaluation results.

Evaluation concept's theoretical foundations and respective suitability

The evaluation concept was primarily based on the Design Science Research Evaluation literature. For each element of the evaluation concept, we tried to utilize the most suitable and adequate components from the literature to ensure a rigorous and scientific evaluation while meeting our evaluation goals.

The fundament of our evaluation concept is the evaluation criteria. To develop them, we mainly aligned with Hevner et al.'s goal-driven framework for developing evaluation criteria [23]. By doing so, it allowed us to create evaluation criteria that captured well our evaluation goals by accounting for the evaluation-related context challenges and incorporating also the actual artifact goals. The latter can be interpreted as a diversion from Hevner et al.'s framework [23] as the framework only accounts for the overall evaluation goals. Therefore by additionally integrating the artifact goals, we could formulate even more precise evaluation criteria, ensuring the generation of more precise evaluation results.

Furthermore, not relying on one evaluation methodology for all evaluation criteria but using different methodologies derived from Design Science Research Evaluation literature for each evaluation criteria cluster, thus effectiveness, supportability, and adoption, allowed us to choose adequate data collection and analysis methods. Hence, we formulated regarding the effectiveness-related criteria, utility hypotheses [35, 37], which according to the literature, can be best assessed with Likert Scale-based surveys [56], while choosing for the supportability-related criteria a more practical evaluation based on Venable et al.'s FEDS framework [46].

Yet, it can be argued that such a plethora of evaluation methodologies and associated data collection and analysis methods might impact the precision and accuracy of the results. Nevertheless, we believe that precisely this diversification makes the evaluation more robust and versatile by ensuring an evaluation accounting for multiple perspectives and being approached by multiple vectors.

The evaluation concept dealing with instantiation context challenges

Aligning with Hevner et al.'s goal-driven framework for developing evaluation criteria [23], we assessed and adjusted our evaluation criteria to the evaluation context challenges. The major challenge was the unavailable direct access to the government's digital infras-

tructure to capture the GaaPIAM's impact directly on the infrastructure. However, using the practitioners as a proxy unveiled itself to be advantageous from a practical accessibility perspective and also accuracy-wise. Particularly regarding the latter, it is more precise and easy to argue a causal effect of the GaaPIAM on the practitioners than attribute any changes in the actual infrastructure to the Method.

Additionally, conducting a pre-interview as well as a post-interview with an as high as possible temporal distance revealed itself to be advantageous for the evaluation. Firstly, although the increased time frame was not long enough to capture significant changes in the infrastructure, it still allowed the practitioners to incorporate some of the insights from the workshops into their work. Secondly, the use of two distinct evaluation events allowed for a before and after instantiation analysis. Consequently, we could better and more rigorously capture changes related to insights generated from the workshops, ultimately improving the evaluation of the supportability-related evaluation criteria.

Additionally, utilizing the practitioners' challenges identified during the pre-interviews served as a good reference point for the post-interview to track how practitioners incorporated insights from the workshop to cope with their respective challenges. Thus, it also benefited the evaluation of the supportability-related evaluation criteria.

The adoption by practitioners evaluation criterion and the use of the Net Promoter Score

One drawback identified during the development of evaluation criteria was the lack of criteria that can adequately capture the adoption willingness of an artifact by practitioners. Although we identified criteria in the literature that might indicate or indirectly lead to the adoption of an artifact by the practitioners, we perceived the lack of an evaluation criterion that can capture this adoption willingness by practitioners precisely. Therefore, we introduced the "adoption by practitioners" as a new evaluation criterion. We defined the criterion as the sustainable and recurring use of an artifact by practitioners. Thus, it aims to evaluate the artifact's overall utility by capturing if practitioners utilize it over time. Inherently this required an extensive evaluation time period. However, in the context of the thesis, only a limited time horizon of approximately three to four weeks was available since the instantiation of the GaaPIAM. This led us to utilize the practitioners' recommendation willingness as a proxy for the adoption. And although the Design Science Research Evaluation theory does not propose suitable data collection methods for capturing the recommendation willingness, the field of behavioral sciences proposes the Net Promoter Score as a well-established measure to assess exactly this recommendation willingness [48, 49]. Although its primary use is in consumer buying behavior analysis [49], it can theoretically be applied to any suitable field.

Nevertheless, the results yielded from applying the Net Promoter Score have to be viewed with caution. First of all, there is no research yet assessing if it can be meaningfully utilized to measure the adoption of an artifact by practitioners. Furthermore, it is not clear how meaningful it is to utilize the Net Promoter Score to evaluate the adoption of an artifact whose instantiation currently only occurs in form of a single event.

Therefore, we believe there is a need to further scientifically investigate the use of the

Net Promoter Score to capture the adoption of an artifact by practitioners, and instead, we suggest observing the artifact's use over time with other data collection methods. Regardless of the question if the Net Promoter Score is an adequate means to evaluate the adoption by practitioners, we perceive the adoption by practitioners as a valuable evaluation criterion for artifacts with the possibility of a more long-term evaluation time frame.

The evaluation's sample size

A major limitation of this thesis at hand is the limited sample size, thus, not permitting the meaningful statistical analysis of the quantitative data generated by the survey's Likert scale questions as well the Net Promoter Score. However, regarding the former, we used a mixed-method approach by asking each practitioner two open questions related to the survey's Likert scale questions during the post-interview. Thus, enabling us to fortify the results from the survey's limited quantitative data with qualitative data to ensure rigor.

9 Conclusion

The evaluation conducted in this thesis represents the evaluation of the third iteration of the GaaPIAM [18]. While the previous evaluation uses a wider range of evaluation criteria with the goal of eradicating teething troubles of the Method and stabilizing its application [21], the present evaluation aims to evaluate if the purpose and goals of the stabilized Method are met.

In general, this can also be affirmed as the previous chapters demonstrate that the GaaPIAM meets its objectives to support practitioners to understand better the GaaP approach, identify gaps and optimization potentials in their infrastructures, and generate concrete actionable proposals for improvement regarding their infrastructures.

9.1 Summary

The thesis at hand is organized along three consecutive research questions, which we answer in the following to conclude the thesis.

Research Question 1: What are the criteria for evaluating the GaaPIAM?

The first research question concerns the identification of evaluation criteria relevant to the evaluation of the GaaPIAM, which then forms the basis of the consecutive evaluation concept. We finally developed a total of nine evaluation criteria where each triple is either related to capturing the artifact's effectiveness, its supportability, and its adoption by practitioners while accounting for the artifact's specific goals.

Research Question 2: How can the GaaPIAM be adequately evaluated?

Building upon the insights of the first research question, we then develop an evaluation concept guiding the evaluation of GaaPIAM. The concept utilizes theoretical concepts and evaluation practices from literature based on suitability to best capture the purpose of each evaluation criterion. The final evaluation concept is fully configured along four stages for each evaluation criterion. Part of the evaluation concept outcome is also the semi-structured interviews, surveys, and field note guidelines, which are based on its configuration. These are utilized during the instantiation of the GaaPIAM to collect and analyze the data.

Research Question 3: How can the evaluation concept be applied in practice?

The last research question concerns the actual application of the evaluation concept during

the instantiation of the GaaPIAM. The results demonstrate that the GaaPIAM supports the practitioners to understand better the GaaP approach, identify gaps and optimization potentials in their infrastructure, and generate concrete actionable proposals for improving their infrastructure.

9.2 Future Work

The GaaPIAM is a stable artifact that demonstrated its utility through the performed evaluation in this thesis.

However, it must be broadly communicated and applied to support governments in transforming their infrastructures following the GaaP approach. In this context, it further requires new instantiation ways that make it more handy, portable, and cost-efficiently reusable. Hence, the development of a reference architecture in form of a checklist might pose a possible future contribution. Furthermore, the GaaPIAM can be utilized to mitigate platform-hindering legislations and regulations by identifying the respective barriers and raising awareness if communicated to the relevant decision-makers.

Regarding the developed evaluation concept, it shows how the use of different evaluation methodologies from literature can lead to a rigorous and versatile evaluation if meaningfully configured. Nevertheless, there is a need for further evaluation of the evaluation criterion "adoption by practitioners" to determine its suitability for evaluating with Design Science Research generated artifacts.

List of Figures

List of Tables

Bibliography

[1] T. O'Reilly. "Government as a Platform". In: *Innovations: Technology, Governance, Globalization* 6.1 (2011), pp. 13–40.

[2] M. Janssen and E. Estevez. "Lean government and platform-based governance—Doing more with less". In: *Government Information Quarterly* 30 (2013), S1–S8.

[3] J. Millard. "ICT-enabled public sector innovation: trends and prospects". In: *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*. 2013, pp. 77–86.

[4] J. Millard. "Open governance systems: Doing more with more". In: *Government Information Quarterly* 35.4 (2018), S77–S87.

[5] S. Sangiambut and R. Sieber. "The Civic Open Data and Crowdsourcing App Ecosystem: Actors, Materials, and Interventions." In: *Journal of the Urban & Regional Information Systems Association* 28.1 (2017).

[6] A. Brown, J. Fishenden, M. Thompson, and W. Venters. "Appraising the impact and role of platform models and Government as a Platform (GaaP) in UK Government public service reform: Towards a Platform Assessment Framework (PAF)". In: *Government Information Quarterly* 34.2 (2017), pp. 167–182.

[7] H. Margetts and A. Naumann. "Government as a platform: What can Estonia show the world". In: *Research paper, University of Oxford* (2017).

[8] A. Cordella and A. Paletti. "Government as a platform, orchestration, and public value creation: The Italian case". In: *Government Information Quarterly* 36.4 (2019), p. 101409.

[9] R. Singh. "Give me a database and I will raise the Nation-State". In: *South Asia: Journal of South Asian Studies* 42.3 (2019), pp. 501–518.

[10] S. Mukhopadhyay, H. Bouwman, and M. P. Jaiswal. "An open platform centric approach for scalable government service delivery to the poor: The Aadhaar case". In: *Government Information Quarterly* 36.3 (2019), pp. 437–448.

[11] E. Styrin, K. Mossberger, and A. Zhulin. "Government as a platform: Intergovernmental participation for public services in the Russian Federation". In: *Government Information Quarterly* 39.1 (2022), p. 101627.

[12] B. Bender and M. Heine. "Government as a Platform? Constitutive Elements of Public Service Platforms". In: *International Conference on Electronic Government and the Information Systems Perspective*. Springer. 2021, pp. 3–20.

[13] J. Sheridan and J. Tennison. "Linking UK government data". In: *Ldow*. 2010.

[14] Y. Akatkin and E. Yasinovskaya. "Data-centricity as the key enabler of digital government: Is Russia ready for digital transformation of public sector". In: *International Conference on Electronic Governance and Open Society: Challenges in Eurasia*. Springer. 2019, pp. 439–454.

[15] D. Rozhkova, N. Rozhkova, and U. Blinova. "Development of the e-Government in the Context of the 2020 Pandemics". In: *International Conference on Advances in Digital Science*. Springer. 2021, pp. 465–476.

[16] E. Zaramenskikh and M. Lyubarskaya. "Integration of digital services within the framework of the implementation of "government as a platform"(GAAP) model on the example of a social fund". In: *Digital Transformation and New Challenges*. Springer, 2020, pp. 1–10.

[17] P. Kuhn, M. Buchinger, D. Balta, and F. Matthes. "Barriers of applying Government as a Platform in Practice: Evidence from Germany". In: *Proceedings of the 55th Hawaii International Conference on System Sciences*. 2022.

[18] P. Kuhn, S. Dallner, M. Buchinger, D. Balta, and F. Matthes. "Toward Government as a Platform: An Analysis Method for Public Sector Infrastructure". In: 2023.

[19] A. R. Hevner, S. T. March, J. Park, and S. Ram. "Design science in information systems research". In: *MIS quarterly* (2004), pp. 75–105.

[20] S. T. March and G. F. Smith. "Design and natural science research on information technology". In: *Decision support systems* 15.4 (1995), pp. 251–266.

[21] P. Kuhn, S. Dallner, M. Buchinger, D. Balta, and F. Matthes. "Toward Government as a Platform: An Analysis Method for Public Sector Infrastructure". In: 2022.

[22] J. Venable, J. Pries-Heje, and R. Baskerville. "A comprehensive framework for evaluation in design science research". In: *International conference on design science research in information systems*. Springer. 2012, pp. 423–438.

[23] A. Hevner, N. Prat, I. Comyn-Wattiau, and J. Akoka. "A pragmatic approach for identifying and managing design science research goals and evaluation criteria". In: *AIS SIGPrag Pre-ICIS workshop on" Practice-based Design and Innovation of Digital Artifacts"*. 2018.

[24] S. H. Schelin. "E-government: an overview". In: *Modern public information technology systems: Issues and challenges* (2007), pp. 110–126.

[25] J. S. Lubbers. "Paperwork Redux: The (Stronger) Paperwork Reduction Act of 1995". In: *Admin. L. Rev.* 49 (1997), p. 111.

[26] H. Seo and S. Myeong. "The priority of factors of building government as a platform with analytic hierarchy process analysis". In: *Sustainability* 12.14 (2020), p. 5615.

[27] D. Jamieson, R. Wilson, and M. Martin. "Is the GaaP wider than we think? Applying a sociotechnical lens to Government-as-a-Platform". In: *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*. 2020, pp. 514–517.

[28] L. Smorgunov. "Government as a Platform: Critics of a Technocratic Culture of Public Governance in Digital Era". In: *International Conference on Digital Transformation and Global Society*. Springer. 2019, pp. 41–54.

[29] B. der Justiz. "Gesetz zur Verbesserung des Onlinezugangs zu Verwaltungsleistungen (Onlinezugangsgesetz-OZG)". In: 16 (2018).

[30] W. Bartels. "Onlinezugangsgesetz (OZG)". In: *Fachlexikon der Sozialen Arbeit*. Nomos Verlagsgesellschaft mbH & Co. KG. 2022, pp. 624–625.

[31] L. Danneels, S. Viaene, and J. Van den Bergh. "Open data platforms: Discussing alternative knowledge epistemologies". In: *Government Information Quarterly* 34.3 (2017), pp. 365–378.

[32] D. Linders. "We-Government: an anatomy of citizen coproduction in the information age". In: *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*. 2011, pp. 167–176.

[33] D. Linders. "From e-government to we-government: Defining a typology for citizen coproduction in the age of social media". In: *Government information quarterly* 29.4 (2012), pp. 446–454.

[34] D. Jones, S. Gregor, et al. "The formulation of an information systems design theory for e-Learning". In: *First International Conference on Design Science Research in Information Systems and Technology, Claremont, CA*. 2006, pp. 356–373.

[35] S. Gregor. "Building theory in the sciences of the artificial". In: *Proceedings of the 4th international conference on design science research in information systems and technology*. 2009, pp. 1–10.

[36] H. A. Simon. "The sciences of the artificial". In: *Cambridge, MA* (1969).

[37] J. Venable. "A framework for design science research activities". In: *Emerging Trends and Challenges in Information Technology Management: Proceedings of the 2006 Information Resource Management Association Conference*. Idea Group Publishing. 2006, pp. 184–187.

[38] S. Aier and C. Fischer. "Criteria of progress for information systems design theories". In: *Information Systems and E-Business Management* 9.1 (2011), pp. 133–172.

[39] U. Frank. *Evaluation von Artefakten in der Wirtschaftsinformatik*. na, 2000.

[40] T. Bucher, C. Riege, and J. Saat. "Evaluation in der gestaltungsorientierten Wirtschaftsinformatik-Systematisierung nach Erkenntnisziel und Gestaltungsziel". In: (2008).

[41] A. Cleven, P. Gubler, and K. M. Hüner. "Design alternatives for the evaluation of design science research artifacts". In: *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*. 2009, pp. 1–8.

[42] M. Rosemann and I. Vessey. "Toward improving the relevance of information systems research to practice: the role of applicability checks". In: *Mis Quarterly* (2008), pp. 1–22.

[43] D. Pfeiffer and B. Niehaves. "Evaluation of conceptual models-a structuralist approach". In: *ECIS 2005 Proceedings* (2005), p. 43.

[44] K. Peffers, M. Rothenberger, T. Tuunanen, and R. Vaezi. "Design Science Research Evaluation". In: vol. 7286. May 2012, pp. 398–410. ISBN: 978-3-642-29862-2. DOI: 10.1007/978-3-642-29863-9_29.

[45] C. Sonnenberg and J. Vom Brocke. "Evaluation patterns for design science research artefacts". In: *European design science symposium*. Springer. 2012, pp. 71–83.

[46] J. Venable, J. Pries-Heje, and R. Baskerville. "FEDS: a framework for evaluation in design science research". In: *European journal of information systems* 25.1 (2016), pp. 77–89.

[47] A. R. Hevner. "A three cycle view of design science research". In: *Scandinavian journal of information systems* 19.2 (2007), p. 4.

[48] F. F. Reichheld. "The one number you need to grow". In: *Harvard business review* 81.12 (2003), pp. 46–55.

[49] F. Reichheld. *The ultimate question 2.0 (revised and expanded edition): How net promoter companies thrive in a customer-driven world*. Harvard Business Review Press, 2011.

[50] J. Corbin and A. Strauss. "Strategies for qualitative data analysis". In: *Basics of Qualitative Research. Techniques and procedures for developing grounded theory* 3.10.4135 (2008), p. 9781452230153.

[51] R. Likert. "A technique for the measurement of attitudes." In: *Archives of psychology* (1932).

[52] W. L. Richman, S. Kiesler, S. Weisband, and F. Drasgow. "A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews." In: *Journal of applied psychology* 84.5 (1999), p. 754.

[53] J. Phillippi and J. Lauderdale. "A guide to field notes for qualitative research: Context and conversation". In: *Qualitative health research* 28.3 (2018), pp. 381–388.

[54] S. G. Kwak and J. H. Kim. "Central limit theorem: the cornerstone of modern statistics". In: *Korean journal of anesthesiology* 70.2 (2017), pp. 144–156.

[55] S. Gregor and A. Hevner. "Positioning and Presenting Design Science Research for Maximum Impact". In: *MIS Quarterly* 37 (June 2013), pp. 337–356. DOI: 10.25300/MISQ/2013/37.2.01.

[56] J. Iivari, M. Hansen, and A. Haj-Bolouri. "A Proposal for Minimum Reusability Evaluation of Design Principles". In: *European Journal of Information Systems* 30 (Jan. 2021), pp. 286–303. DOI: 10.1080/0960085X.2020.1793697.

[57] J. W. Creswell. "Mixed-method research: Introduction and application". In: *Handbook of educational policy*. Elsevier, 1999, pp. 455–472.