

The Evolution of Platform Engineering in the Public Sector: A Longitudinal Study of Adoption, Maturity, and Motivations (2024–2026)

Hans Kristian Flaatten
Norwegian Labour and Welfare Administration (Nav)
hans.kristian.flaatten@nav.no

Abstract

Context

Platform Engineering and Internal Developer Platforms (IDPs) have become a leading approach to scaling software delivery, yet longitudinal empirical data on how adoption, motivations, and maturity evolve remains scarce.

Objective

This study tracks the evolution of platform engineering adoption, tooling, maturity, and motivations in the Norwegian public sector over two years (2024–2026).

Methods

We conducted a repeated cross-sectional survey with 35 organizations in 2024 (collected Q4 2023) and 45 in 2026 (collected Q4 2025 – Q1 2026), with 60% organizational overlap enabling panel analysis. We apply Fisher's exact tests to cross-wave comparisons and McNemar's exact test to paired panel comparisons.

Results

IDP adoption reached 92% (n=48) and Kubernetes 83% (n=48) by 2026. Among the 48 platform entries (45 organizations, three of which run two platforms), developer agility remains the primary motivation (91%, n=44), but secondary drivers shifted: security rose 25 percentage points (pp) to 82% (p=0.034, uncorrected), cost optimization rose 15pp to 52% (p=0.237), and recruitment declined 11pp to 36% (p=0.471). Among the motivations, only the security increase reaches $\alpha=0.05$ at the individual test level; it does not survive Bonferroni correction. In the paired panel (24 organizations with a platform in both waves), 9 organizations added security as a motivation and 1 dropped it (McNemar p=0.021). Using the CNCF Platform Maturity Model, we find that modal operational maturity shifted from Level 2 to Level 3, while measurement maturity remains at Level 1, with 49% of platform entries (n=43) reporting ad hoc practices.

Conclusion

Platform engineering in the Norwegian public sector is transitioning from experimental initiative toward standardized infrastructure. Organizational emphasis is shifting from agility and recruitment toward security and cost optimization, though small sample sizes limit statistical power. The persistent gap between operational and measurement maturity poses a risk for teams expected to demonstrate return on investment.

Keywords: platform engineering, internal developer platform, DevOps, cloud-native, longitudinal study, public sector, Kubernetes

1. Introduction

How organizations deliver and maintain software is changing. Driven by the growing complexity of cloud-native architectures and the resulting cognitive load on development teams [1], organizations

are adopting Platform Engineering [25]. By establishing dedicated platform teams to build Internal Developer Platforms (IDPs), organizations aim to abstract away infrastructure complexity, providing application developers with self-service capabilities and paved roads that enforce compliance and best practices by default.

While anecdotal evidence and industry reports promote platform engineering, the academic and empirical literature lacks longitudinal data [21]. Much of the existing research focuses on point-in-time snapshots or relies heavily on data from hyper-scale technology companies, whose operational contexts differ substantially from enterprise or public sector environments. Consequently, it is difficult to determine whether the stated motivations for adopting IDPs, or the maturity models used to evaluate them, hold true as these initiatives scale and mature over multiple years.

To address this gap, this study uses the Norwegian public sector as a microcosm for enterprise IT evolution. The Norwegian public sector is highly digitized, heavily funded, and characterized by a strong culture of early technological adoption and knowledge sharing across agency boundaries. This environment provides a bounded context for studying how platform engineering matures in practice.

We present a repeated cross-sectional analysis based on survey data collected across two distinct periods: Q4 2023 (establishing the 2024 baseline) and Q4 2025 – Q1 2026. By analyzing responses from 35 organizations in the first wave and 45 in the second—with 27 organizations (60%) participating in both waves—we aim to empirically map the trajectory of platform engineering. The 2026 wave alone represents approximately 240,000 employees across 45 unique organizations, spanning state agencies, state-owned enterprises, and municipalities.

This paper addresses the following Research Questions (RQs):

- **RQ1 (Adoption & Scale):** How has the adoption and organizational scale of Internal Developer Platforms evolved over the two-year period?
- **RQ2 (Motivations):** What are the primary organizational drivers for building IDPs, and how have these motivations shifted over the two-year period?
- **RQ3 (Maturity):** How is platform maturity evolving in practice, specifically when evaluated against the multidimensional CNCF Platform Maturity Model?
- **RQ4 (Technology & Tooling):** What technology choices and toolchains dominate modern platform stacks, and how are they consolidating?

Our findings show that platform engineering has become the dominant operating model in this community. We document a shift in organizational motivation—from talent acquisition toward security enforcement and cost optimization—and identify a growing gap between operational maturity and the ability to quantify platform value.

The remainder of this paper is structured as follows: Section 2 outlines the background and related work. Section 3 details the survey methodology and data normalization processes. Section 4 presents the empirical results across our four research questions. Section 5 discusses the implications of these findings. Section 6 addresses threats to validity, and Section 7 concludes with directions for future research.

2. Background and Related Work

2.1. DevOps, Cognitive Load, and the Rise of Platform Engineering

The original premise of DevOps—“you build it, you run it”—sought to eliminate silos between software development and IT operations. However, as cloud-native systems (e.g., Kubernetes, service meshes, distributed tracing) grew in complexity, this approach inadvertently shifted a large cognitive burden onto application developers. Skelton and Pais [1], in *Team Topologies*, identified this cognitive overload as a primary bottleneck to software delivery performance. To alleviate this, they proposed the “Platform Team,” a specialized grouping dedicated to providing an internal product—an Internal Developer

Platform (IDP)—that abstracts away infrastructure complexity, allowing stream-aligned teams to focus on delivering end-user functionality.

The empirical link between DevOps capabilities and organizational performance was established by Forsgren, Humble, and Kim [2] through the DORA (DevOps Research and Assessment) research program. Their work identified key capabilities—including continuous delivery, infrastructure automation, and monitoring—that predict software delivery performance and organizational outcomes. The annual DORA State of DevOps reports [3] have since tracked these capabilities globally, providing a benchmark against which domain-specific studies such as ours can be compared. Our study differs from DORA in its focus on the *organizational* unit of analysis (the platform initiative) rather than individual team capabilities, and in its restriction to a single national public sector context.

2.2. The CNCF Platform Engineering Maturity Model

To standardize the evaluation of platform engineering initiatives, the Cloud Native Computing Foundation (CNCF) Technical Advisory Group (TAG) for App Delivery published the *Platform Engineering Maturity Model* [4]. This model provides a framework for organizations to assess their platform’s evolution across five distinct aspects, measured across four levels of maturity (Level 1: Provisional, Level 2: Operationalized, Level 3: Scalable, Level 4: Optimizing).

The five aspects evaluated in our survey are:

1. **Investment:** How staff and funding are allocated. Maturity ranges from voluntary, temporary “tiger teams” (L1) to dedicated cost centers (L2), treating the platform “as a product” (L3), and finally, an enabled ecosystem (L4).
2. **Adoption:** How users discover and utilize the platform. This evolves from erratic discovery (L1), to extrinsic mandates (L2), to intrinsic pull where users choose the platform for its value (L3), and participatory contribution (L4).
3. **Interfaces:** How capabilities are consumed. Moving from custom, manual processes (L1), to standard paved roads (L2), to self-service solutions (L3), and fully integrated workflows embedded in the developer experience (L4).
4. **Operations:** How platforms are maintained. Progressing from reactive, request-driven development (L1), to centrally tracked services (L2), centrally enabled automated lifecycles (L3), and fully managed services with shared responsibility models (L4).
5. **Measurement:** How feedback is gathered and utilized. Starting with ad hoc, anecdotal decisions (L1), moving to consistent collection of metrics (L2), data yielding strategic insights (L3), and a culture of continuous quantitative and qualitative improvement (L4).

The measurement aspect of this model connects to a broader tradition of software metrics research, particularly the Goal-Question-Metric (GQM) paradigm introduced by Basili, Caldiera, and Rombach [5], which provides a structured approach to defining and collecting metrics aligned with organizational goals. Our finding that measurement maturity trails other aspects may partly reflect the absence of such structured frameworks in platform engineering practice.

2.3. Industry Surveys and Benchmarks

Several industry surveys have tracked platform engineering adoption in recent years. The Humanitec State of Platform Engineering reports [6] survey a global, predominantly private-sector audience and have documented similar adoption trajectories for Kubernetes, Terraform, and ArgoCD, though with different maturity distributions reflecting a more heterogeneous sample. The Puppet State of Platform Engineering report [7] similarly found that organizations with more mature platform practices report higher developer satisfaction and deployment frequency. The CNCF End-User Technology Radar [8] provides periodic snapshots of technology adoption among CNCF end-user organizations, offering an external reference point for our tooling convergence findings.

Our study complements these industry surveys in two ways: (1) by focusing on a bounded, well-characterized population (the Norwegian public sector) rather than a self-selected global sample, enabling more controlled cross-wave comparison; and (2) by applying the CNCF Maturity Model systematically across two time points, providing longitudinal maturity trajectory data that industry surveys, typically point-in-time snapshots, do not offer. Where possible, we note points of comparison with these surveys in Section 5. However, direct numerical comparison is limited by differences in sampling methodology, population characteristics, and survey instruments.

2.4. Platform Engineering in the Public Sector

The public sector operates under different constraints than private industry [26]. Organizations face rigid procurement laws, strict regulatory requirements, and elevated demands for security and data sovereignty. Conversely, in Norway, agencies such as Nav (the Norwegian Labour and Welfare Administration) and Skatteetaten (the Norwegian Tax Administration) operate with large budgets and scale, often pioneering cloud-native adoption before their private-sector counterparts [17][18]. Norway consistently ranks among the top countries in digital government maturity [24], providing a context where public-sector organizations have both the mandate and capacity for technological innovation. Understanding platform engineering in this context offers practical insight into how large, regulated enterprises reconcile agility with compliance.

2.5. Theoretical Framing

This study draws on two complementary theoretical lenses to interpret adoption and convergence patterns.

First, Rogers’ *Diffusion of Innovations* theory [9] provides a framework for understanding the adoption lifecycle. With 92% of the surveyed community reporting IDPs, the population appears to have passed the “early majority” threshold and entered the “late majority” stage, where remaining non-adopters face structural barriers rather than awareness gaps. This framing has implications for how we interpret the motivation shifts observed in Section 4.2 — as adoption saturates, organizational discourse shifts from *whether* to adopt platforms to *what those platforms should prioritize*.

Second, DiMaggio and Powell’s *institutional isomorphism* [10] offers an explanation for the technology convergence observed in RQ4. Their three mechanisms — mimetic (imitation of successful peers), normative (professionalization through shared networks), and coercive (regulatory or procurement pressure) — may explain why 45 independently governed organizations converge on a narrow toolset rather than exhibiting the diversity one would expect from independent technology evaluations.

These theoretical perspectives are not tested as formal hypotheses in this study. Rather, they serve as interpretive lenses that inform our discussion of the empirical findings (Section 5). The study’s primary contribution is descriptive-exploratory: characterizing the state and trajectory of platform engineering in a bounded population where little prior empirical data exists [20][21].

3. Methodology

3.1. Research Design and Data Collection

This study employs a repeated cross-sectional survey design with a longitudinal subset. We adopt a pragmatist worldview [19] and position the study as descriptive-exploratory: we aim to characterize the state and trajectory of platform engineering adoption rather than test pre-specified hypotheses. This design is consistent with established precedent for survey research in software engineering where population characteristics are poorly understood [20][21] and follows the quality guidelines of Molléri et al. [11][22] and the guidance on questionnaire design of Kitchenham and Pfleeger [12].

Data was collected at two distinct points in time: the first wave in Q4 2023 (November–December 2023, establishing the 2024 baseline) and the second wave spanning Q4 2025 to Q1 2026 (December 2025 to February 2026). The survey was distributed primarily through *Offentlig PAAS* (Public PaaS), an active

cross-agency community of practice dedicated to platform engineering within the Norwegian public sector.

Survey Instrument and Design: The survey instrument was developed by the author based on prior community discussions and the CNCF Platform Maturity Model [4]. It was not based on a previously validated instrument, as no suitable validated instrument for platform engineering maturity exists in the literature. The 2024 instrument comprised 33 items in three sections: (1) general platform engineering (17 items: organization, organizational type and size, respondent role, platform existence and name, motivations, self-service capabilities, number of application teams, programming languages, and five CNCF maturity items, one per aspect with four response options corresponding to Levels 1–4), (2) Kubernetes (5 items: use, distributions, years in production, cluster count, application count), and (3) cloud (11 items: cloud use, primary provider, multi-cloud, years in cloud, serverless, and tooling for monitoring, packaging, deployment, network security, application security, and infrastructure automation). Most items are closed (single-select, multi-select, or predefined ranges), but the organization and platform name items are open text, and most multi-select items accept an “Other” write-in, which respondents used. The 2026 instrument kept the 2024 items and added two Kubernetes items: whether the organization develops its own operators, and whether it runs databases on Kubernetes (35 items in total). One response scale changed: the top cluster-count option “more than 10” (2024) was split into 10–19, 20–49 and 50+ (2026); we merge these into “>10” for comparison. The full instrument, with the introduction text shown to respondents, is provided as supplementary material.

Recruitment and Response: In both waves, invitations were posted to the Offentlig PAAS Slack workspace (about 1,800 members at the time of the 2026 wave) and shared on LinkedIn. One reminder was sent via Slack approximately two weeks after the initial invitation in each wave. According to the community’s member records, it had 84 member organizations at the time of the 2026 wave, of which 45 responded (54% organizational response rate). Membership at the time of the 2024 wave was not recorded, so no response rate is reported for that wave. No individual incentives were offered; participation was framed as contributing to community knowledge.

Pre-normalization Respondent Characteristics: The 2026 wave received 73 individual form submissions from 45 unique organizations. After normalization (Section 3.3), these were reduced to 48 platform entries. Respondent roles prior to normalization were: Platform Developer (38%), Tech Lead (17%), Manager (15%), Architect (11%), DevOps (7%), Operations (4%), Application Developer (3%), and Other (5%). The 2024 wave received 35 submissions corresponding to 35 unique organizations (no duplicates were received). Roles were: Platform Developer (40%), Architect (17%), Tech Lead (11%), Operations (9%), Manager (9%), DevOps (6%), and other roles (9%).

Of the 45 organizations in the 2026 wave, 27 (60%) also participated in the 2024 baseline, providing a substantial panel subset for within-cohort comparisons. The remaining 18 organizations are new to the 2026 wave, while 8 organizations from 2024 did not participate in 2026. To address potential cohort composition effects, we report both full-sample comparisons and, for key findings, sensitivity analyses restricted to the 27-organization panel subset.

Given the bounded nature of the community and the organizational level of the data, this study relies primarily on descriptive statistics to characterize macro-trends. Given the exploratory nature of this study and small sample sizes, we report uncorrected p-values from Fisher’s exact test for all key percentage-point changes and interpret them as indicators of effect strength rather than confirmatory hypothesis tests. Where multiple comparisons are performed (e.g., the seven motivations in Table 1), we note the Bonferroni-corrected threshold but do not apply it as a strict acceptance criterion. For paired comparisons within the panel, where the same organization is observed in both waves, we use McNemar’s exact test [27] on the discordant pairs. Panel pairs use one platform entry per organization: for dual-platform organizations, the first-listed 2026 entry. Using the second-listed entry instead gives, for security, 8 gained and 1 lost ($p=0.039$) and, for cost, 6 gained and 2 lost; conclusions are unchanged.

All percentages are rounded to the nearest integer (halves rounded up, e.g. $21/40 = 52.5\% \rightarrow 53\%$) unless otherwise noted, and differences (Δ , pp) are differences of the rounded percentages; where this introduces $\pm 0.5\text{pp}$ ambiguity (e.g., $17/30 = 56.7\% \approx 57\%$), the raw counts are provided alongside.

3.2. Demographics and Unit of Analysis

The dataset captures a broad cross-section of Norwegian public-sector organizations engaged in platform engineering:

- **2024 Baseline:** 35 unique organizations, representing approximately 177,000 employees. Organizational budgets range from under 500 MNOK (small directorates) to over 100,000 MNOK (large state agencies and state-owned enterprises).
- **2026 Dataset:** 45 unique organizations (48 platform entries), reflecting a 29% increase in organizational participation, representing approximately 240,000 unique employees. Budget data (available for 31 of 45 organizations) shows a similar distribution, with two organizations exceeding 100,000 MNOK. By platform entry, the respondents are predominantly State Agencies (*Etater*, 60%, 29/48), State-Owned Enterprises (*Statsforetak*, 25%, 12/48), alongside Municipalities (*Kommuner*, 8%) and Academic Institutions (4%). For inclusion in the study, “public sector” is defined as organizations in which the Norwegian state holds a majority ownership stake ($\geq 51\%$), including state agencies, state-owned enterprises, and municipally governed entities.

The employee totals ($\sim 177,000$ and $\sim 240,000$) are sums over a per-organization table of public figures, rounded; we do not release that table because it would list the participating organizations, and the published aggregate tables give headcount only in ranges. Organizational headcount and budget figures are sourced from publicly available data (annual reports, government budget proposals, and organizational websites) and are reported per unique organization to avoid double-counting where an organization operates multiple platforms. Norwegian public sector organizations are subject to transparency requirements and publish detailed annual reports, making these figures readily available and of consistent quality. Where respondent organizations are IT platform teams serving a larger parent entity (e.g., a shared IT provider serving a regional health authority), headcount and budget figures reflect the parent organization, since the platform serves the broader workforce. All figures reflect total organizational scale, not IT-specific spending. Headcount data is available for 43 of 48 platform entries in 2026 and all 35 in 2024; budget data is available for 34 of 48 entries in 2026 and 34 of 35 in 2024, with all gaps concentrated among new entrants for which public financial data was not separately published.

The unit of analysis is the *platform entry*. Where an organization operates multiple distinct platforms (e.g., a legacy Docker Swarm platform alongside a modern Kubernetes platform), each is represented as a separate row. This yields 35 platform entries for 2024 and 48 platform entries for 2026 (45 unique organizations, three of which maintain two distinct platforms each). Unless otherwise noted, percentages are computed over platform entries, with the base (n) stated explicitly.

The relationship between data collection and unit of analysis is as follows: data is collected from *individual respondents* (one form submission per person), normalized to *platform entries* (one row per distinct platform), and interpreted at the *organizational level* (what the organization does). For 42 of 45 organizations in 2026, the mapping is one-to-one (one organization = one platform = one row). For three organizations, the mapping is one-to-two (one organization = two distinct platforms = two rows). For analyses where organizational independence matters (e.g., motivations, which reflect organizational strategy), we note that the three dual-platform organizations introduce minor non-independence: their two platform entries may share governance, funding, or strategic priorities. This affects 6 of 48 entries (12.5%). We assess this as unlikely to materially affect results but note it as a limitation (Section 6). Self-reported data (maturity levels, platform scale, motivations) were not independently validated; Section 6 discusses this limitation.

3.3. Data Normalization and Threat Mitigation

A common threat to validity in self-reported organizational surveys is the presence of multiple, potentially conflicting responses from different individuals within the same organization. In the 2026 dataset, 13 organizations submitted a total of 25 duplicate responses. The normalization rules described below were defined prior to examining the individual response data, based on the same principles applied during the 2024 baseline analysis.

To ensure data integrity and prevent highly vocal organizations from skewing the results, we employed a strict normalization process to flatten duplicates into a single canonical row per distinct platform.

1. **Platform Distinction:** Where an organization explicitly operated distinct platforms (e.g., a legacy Docker Swarm platform and a modern Kubernetes platform), these were preserved as separate entries.
2. **Role Preference Heuristics:** For conflicting single-value fields (e.g., maturity level assessments), we prioritized responses based on technical proximity to the platform. In-house technical roles (e.g., Platform Developer, DevOps) were preferred over management or external consultants.
3. **Data Unioning:** For multi-select fields (e.g., programming languages, tooling), we computed the union of all unique values reported by respondents from the same platform.
4. **Conservative Aggregation:** For numerical ranges (e.g., number of application teams, years in production), the highest reported concrete value was selected to reflect the platform's maximum realized scale.

This normalization ensures that each platform is represented equitably, reflecting organizational reality without inflation.

Coding of write-in answers. Write-in text in “Other” fields was counted only when it matched a predefined option exactly; we did not recode free text into predefined categories. Tool mentions are counted per question (a tool named under two questions counts once in each). Where a multi-select answer names several tools, every named tool is counted. Three coding decisions affect reported figures, and we give the alternative result for each: (1) one 2026 write-in naming vulnerability scanning was not counted as the predefined security-scanning self-service option (counting it: 77%, 34/44, $p=0.127$, instead of 75%, 33/44, $p=0.206$); (2) one 2026 answer to the databases-on-Kubernetes question that begins with “No” but then names two data stores was counted as running databases (13/40 running; coding it as “No” instead gives 60%, 24/40, not running and 12/40 running); (3) the database-operator counts in Section 4.4 include every mention in answers reporting databases on Kubernetes, also where several operators are named, but exclude mentions in answers describing proof-of-concept work (including it, CloudNativePG rises from 6 to 7).

3.4. Ethical Considerations

Participation was voluntary and no incentives were offered. The survey asked about organizational practices and technology choices, not about personal opinions or special categories of personal data. The organization name was required, solely to detect and merge duplicate responses; an e-mail address was optional, for follow-up contact, and is excluded from all analysis and published material. The wave 2 introduction read (translated from Norwegian): “Follow-up to the 2023 survey on adoption of application platforms in the public sector, organized by Offentlig PaaS. Aggregated results will be published on the Offentlig PaaS Slack and offentlig-paas.no. All questions are answered with regard to the application platform of your organization and take about 10 minutes to complete. For more information about the survey, contact kontakt@offentlig-paas.no.” The wave 1 introduction was similar but its exact wording was not preserved. Neither text met the information requirements of the Norwegian research ethics guidelines (NESH, §15) [28] or Article 13 of the GDPR: they did not name the data controller or the legal basis, describe what was collected and how long it would be kept, inform respondents of their rights, or say that results would also be published in academic venues. No notification to or assessment by Sikt (the Norwegian Agency for Shared Services in Education and

Research) was sought for either wave. Because respondents were told only that aggregated results would be published, this paper reports organizations only in aggregate, and we publish aggregate frequency tables rather than response-level data (see Data Availability). Future waves will use a revised introduction that meets these requirements; it is included with the instrument in the supplementary material.

4. Results

4.1. RQ1: Adoption and Scale of Internal Developer Platforms

The data from 2024 to 2026 indicates that Internal Developer Platforms have moved from majority adoption to widespread use within the surveyed community. In 2024, 86% of platform entries reported having an IDP (30/35); by 2026, this figure rose to 92% (44/48). While this increase of 6pp is not statistically significant ($p=0.483$, Fisher’s exact test), this is consistent with both figures already representing high baselines.

A parallel trend is the consolidation around Kubernetes as the foundational orchestration layer. Kubernetes usage increased from 66% (23/35) in 2024 to 83% (40/48) in 2026, a 17pp increase that approaches but does not reach conventional significance ($p=0.074$).

Beyond binary adoption, the scale at which these platforms operate has grown. The proportion of platforms serving more than 100 application teams increased from 10% (3/30) in 2024 to 18% (8/44) in 2026. Conversely, the “middle tier” of platforms (serving 11–50 teams) contracted from 40% (12/30) to 30% (13/44), suggesting that some mid-sized platform initiatives may have scaled up to enterprise-wide adoption. “Application team count” serves as a proxy for organizational scale; while absolute developer headcount was not captured, the team metric provides a practical indicator of internal platform penetration.

Four Kubernetes and cloud items were asked in both waves and allow further comparison. Among Kubernetes-using entries, the share reporting more than 10 clusters fell from 35% (8/23) to 24% (9/38, excluding two “don’t know”; $p=0.388$), and the share running more than 500 applications on Kubernetes was unchanged at 32% (7/22 → 12/38; $p=1.000$). The share that has run Kubernetes in production for five years or more rose from 52% (12/23) to 70% (28/40; $p=0.183$; median 5 → 6 years), as expected for a largely returning population. Among entries with a primary cloud provider, 63% (15/24) and 59% (20/34) named at least one serverless service ($p=1.000$); the data record only named services, so a blank answer cannot be told apart from “no” or a skipped question. None of these changes reaches $\alpha=0.05$.

4.2. RQ2: The Evolution of Primary Motivations

To understand the organizational mandates behind IDPs, respondents with an active platform were asked to identify their primary motivations (multiple selections allowed). The cross-wave comparison reveals several directional shifts in why public sector entities invest in platform engineering.

While “Faster application development” (*Raskere applikasjonsutvikling*, Agility) remains the most cited motivation—reported by 93% (28/30) in 2024 and 91% (40/44) in 2026—the secondary drivers show a realignment (see Table 1 and Figure 1).

Table 1: Evolution of Strategic Motivations (2024–2026)

Motivation	2024 (n=30)	2026 (n=44)	Delta (pp)	Fisher p
Agility (Raskere utvikling)	93%	91%	-2	1.000
Quality (Stabilitet og kvalitet)	73%	82%	+9	0.404
Security (Sikkerhet)	57%	82%	+25	0.034
Resources (Ressursoptimalisering)	57%	61%	+4	0.810
Cost (Kostoptimalisering)	37%	52%	+15	0.237
Recruiting (Attraktiv arbeidsplass)	47%	36%	-11	0.471
Tech Diversity (Støtte teknologier)	17%	23%	+6	0.571

Percentages reflect the share of platform-having entries citing each motivation. Fisher's exact test (two-sided) applied to each 2x2 comparison. Given 7 simultaneous tests, a Bonferroni-corrected threshold of $\alpha=0.007$ would be conservative; we report uncorrected p-values and note that only the security increase reaches $\alpha=0.05$.

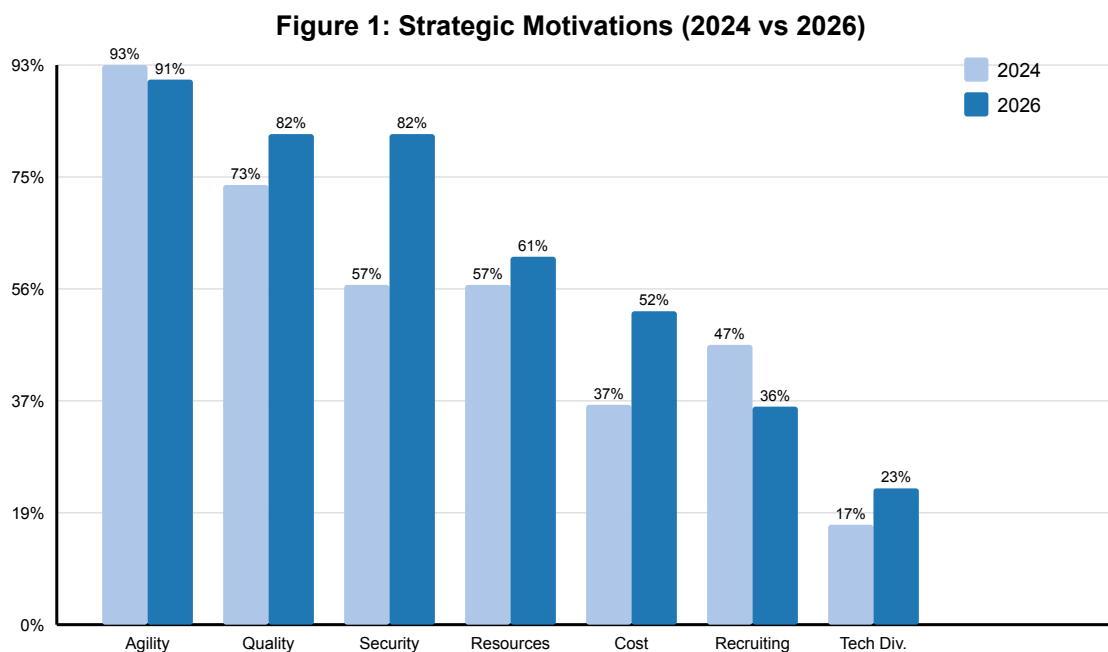


Figure 1: Strategic Motivations for Platform Engineering (2024 vs 2026)

The Rise of Security as a Primary Driver: The largest observed change is the increase in security as a primary driver. In 2024, “Security” (*Sikkerhet*) was a middle-tier motivation at 57% (17/30). By 2026, it rose 25pp to 82% (36/44), tying for the second most cited motivation alongside “Stability and Quality.” This is the only motivation change that reaches $\alpha=0.05$ at the individual test level (uncorrected; see Section 6 for multiple comparisons discussion; OR=3.4, the odds of citing security in 2026 relative to 2024). The trend holds within the panel: of the 24 organizations with a platform in both waves, 13 cited security in 2024 and 21 in 2026; 9 added it and 1 dropped it (McNemar exact $p=0.021$). “Cost optimization” (*Kostoptimalisering*) also rose from 37% to 52% (+15pp), though this increase does not reach statistical significance ($p=0.237$). As cloud infrastructure footprints grow, platforms may increasingly be tasked with enforcing governance, standardizing security boundaries, and managing decentralized spending.

The Decline of the Recruiting Premium: Using modern platforms as a tool for talent acquisition and retention (“Being an attractive workplace,” *Være attraktiv arbeidsplass for utviklere*) declined from 47% (14/30) to 36% (16/44), a decrease of 11pp. However, this change does not reach statistical significance ($p=0.471$), and multiple explanations are plausible, including shifts in the broader technology labor market, a maturing view of platforms as operational necessities rather than recruitment differentiators, or sampling variation.

4.3. RQ3: The Maturity Paradox (CNCF Model)

Organizations with platforms self-assessed their initiatives against the five aspects of the CNCF Platform Maturity Model. Table 2 presents the full cross-wave comparison across all five aspects. The data reveals a divergence: operational capabilities show directional improvement, while measurement practices remain underdeveloped.

Table 2: CNCF Platform Maturity Model — Full Cross-Wave Comparison

Aspect	Level	2024	2026	Δ (pp)
Investment (n=30 / 43)	L1: Provisional	13% (4)	0% (0)	-13
	L2: Operationalized	57% (17)	67% (29)	+10
	L3: Scalable	20% (6)	28% (12)	+8
	L4: Optimizing	10% (3)	5% (2)	-5
Adoption (n=30 / 42)	L1: Provisional	10% (3)	14% (6)	+4
	L2: Operationalized	33% (10)	31% (13)	-2
	L3: Scalable	43% (13)	52% (22)	+9
	L4: Optimizing	13% (4)	2% (1)	-11
Interfaces (n=30 / 43)	L1: Provisional	17% (5)	12% (5)	-5
	L2: Operationalized	33% (10)	30% (13)	-3
	L3: Scalable	37% (11)	49% (21)	+12
	L4: Optimizing	13% (4)	9% (4)	-4
Operations (n=30 / 43)	L1: Provisional	13% (4)	19% (8)	+6
	L2: Operationalized	50% (15)	23% (10)	-27
	L3: Scalable	23% (7)	47% (20)	+24
	L4: Optimizing	13% (4)	12% (5)	-1
Measurement (n=30 / 43)	L1: Provisional	53% (16)	49% (21)	-4
	L2: Operationalized	43% (13)	37% (16)	-6
	L3: Scalable	0% (0)	12% (5)	+12
	L4: Optimizing	3% (1)	2% (1)	-1

Base sizes vary slightly across aspects due to missing responses (1–2 entries with blank maturity fields in 2026).

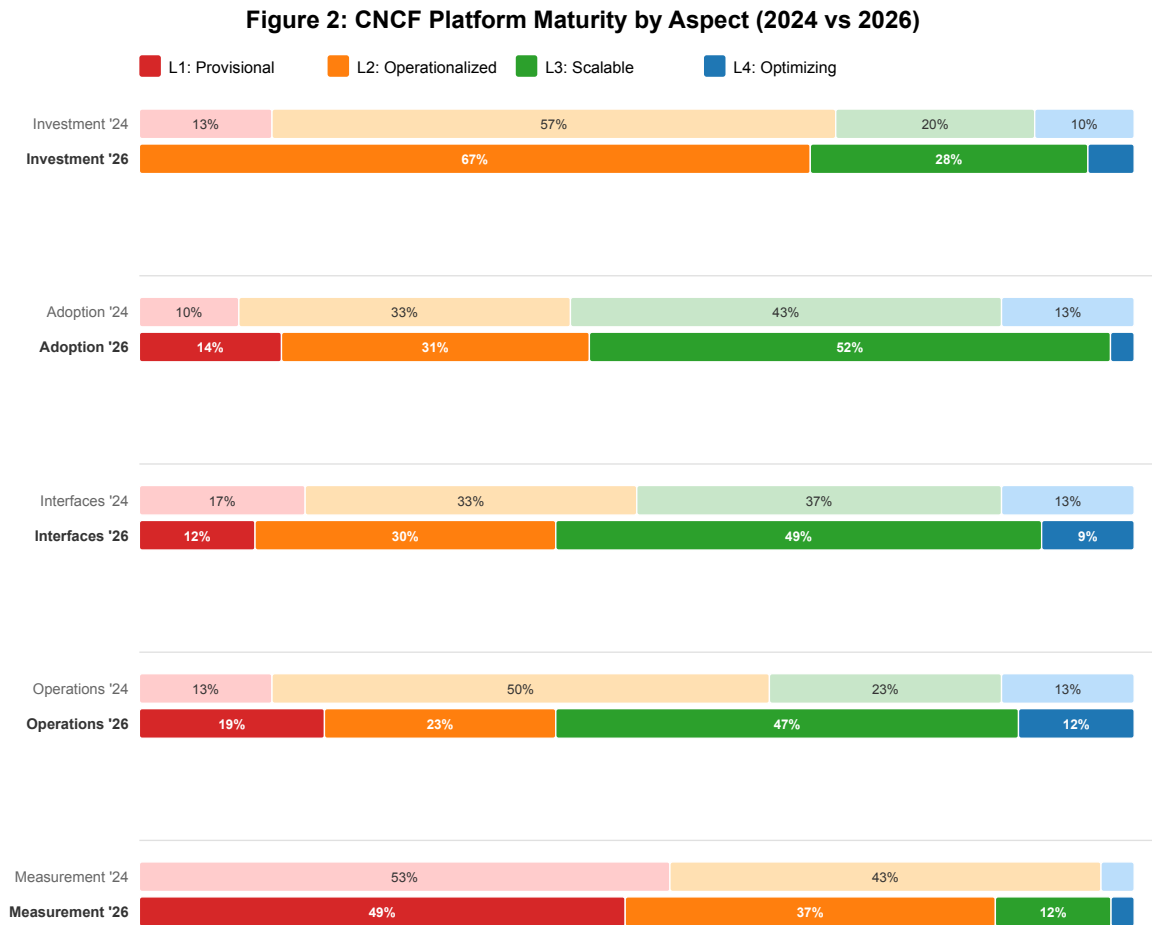


Figure 2: Platform Maturity by Aspect (2024 vs 2026)

- **Investment (Consolidation at L2):** The Provisional (Level 1) phase for Investment was eliminated, dropping from 13% (4/30) to 0% (0/43). The majority consolidated at Level 2 (Dedicated Teams/Cost Centers, 67%, 29/43), while Level 3 (Product-minded) grew from 20% to 28%.
- **Adoption (Growth in L3, Decline in L4):** Adoption at Level 3 (Intrinsic Pull) increased from 43% (13/30) to 52% (22/42), becoming the clear modal response. Level 4 (Participatory) dropped from 13% (4/30) to 2% (1/42), suggesting that while more organizations attract users organically, few have achieved full participatory contribution.
- **Interfaces (Growth in Self-Service):** Level 3 (Self-service solutions) grew from 37% to 49% (21/43), becoming the modal response. The shift comes primarily from L1 and L2, suggesting a broad move toward automated, on-demand interfaces.
- **Operations (Directional Shift with Caveats):** The largest directional change occurred in Operations. In 2024, the modal level was Level 2 (Centrally Tracked, 50%, 15/30). By 2026, the mode shifted to Level 3 (Centrally Enabled, 47%, 20/43). The proportion of entries at Level 3 or above rose from 37% (11/30) to 58% (25/43), though this change does not reach conventional significance ($p=0.097$, Fisher's exact test). However, Level 1 also increased from 13% (4/30) to 19% (8/43). Of the eight entries at Level 1 in 2026, three are new entrants, three returned from Level 2 in 2024, and two remained at Level 1 across both waves. This L1 increase may partly reflect differing interpretations of the

CNCF operational maturity definitions between waves, a concern noted in the threats to validity (Section 6).

- **Measurement (The Persistent Capability Gap):** Measurement remains the weakest aspect. In 2026, 49% of platform entries (21/43) remain at Level 1 (Ad hoc), compared to 53% (16/30) in 2024—a minimal change ($p=0.813$). However, a small vanguard has emerged at Level 3 (Data-driven strategic insights): 12% (5/43) in 2026, up from 0% (0/30) in 2024.

Within-Subject Panel Analysis: To complement the cross-sectional comparison, we analyzed the 24 panel organizations with valid maturity data in both waves (Table 3). This within-subject analysis tracks whether individual organizations moved up, remained stable, or regressed on each maturity aspect.

Table 3: Panel Maturity Transitions (n=24 organizations, 2024→2026)

Aspect	Improved	Stable	Regressed	Sign test p
Investment	3 (12%)	18 (75%)	3 (12%)	1.000
Adoption	5 (21%)	11 (46%)	8 (33%)	0.581
Interfaces	9 (38%)	5 (21%)	10 (42%)	1.000
Operations	8 (33%)	8 (33%)	8 (33%)	1.000
Measurement	4 (17%)	16 (67%)	4 (17%)	1.000
All	29 (24%)	58 (48%)	33 (28%)	—

Sign test: two-sided exact binomial, $H_0: P(\text{improved}) = P(\text{regressed})$. No aspect shows a statistically significant directional trend. For the two dual-platform panel organizations, the first-listed 2026 entry is paired with the 2024 entry; pairing with the higher-rated 2026 entry instead gives 33 improved, 55 stable, and 32 regressed, with the same conclusion.

The panel analysis reveals substantial individual-level movement but no statistically significant aggregate directional trend. Approximately half (52%) of organization-aspect pairs changed level between waves, split roughly evenly between improvement and regression. Notable patterns include: (1) Investment is the most stable aspect (75% unchanged), consistent with funding decisions being slow-moving organizational commitments; (2) Interfaces shows the most turbulence (only 21% stable), suggesting active experimentation in how platforms expose capabilities; (3) Operations gravitates toward Level 3, with 6 of 13 organizations at Level 2 in 2024 advancing to Level 3 by 2026; (4) all four organizations at Adoption Level 4 in 2024 moved down by 2026 (to L3 or below), possibly reflecting more critical self-assessment rather than actual regression. The absence of aggregate directional trends, despite substantial individual movement, suggests that the sector exhibits high variance in maturity trajectories rather than uniform upward progress.

4.4. RQ4: Convergence of the Toolchain

The technology stack of public sector platforms is converging on managed Kubernetes, declarative infrastructure, and GitOps-based deployment. This convergence has consolidated further between 2024 and 2026 (see Tables 4–6 and Figure 3).

Table 4: Kubernetes Distribution Adoption

Distribution	2024 (n=22)	2026 (n=40)	Δ (pp)
AKS	45% (10/22)	40% (16/40)	-5
OpenShift	23% (5/22)	22% (9/40)	-1
GKE	18% (4/22)	20% (8/40)	+2
Vanilla K8s	23% (5/22)	12% (5/40)	-11
Tanzu	14% (3/22)	8% (3/40)	-6
EKS	5% (1/22)	8% (3/40)	+3
Talos	—	8% (3/40)	+8
Nutanix K8s Engine	5% (1/22)	5% (2/40)	0
ARO (Azure Red Hat OpenShift)	—	5% (2/40)	+5

Multi-select: organizations may report multiple distributions. Base is the number of platform entries reporting at least one Kubernetes distribution (n=22 in 2024, n=40 in 2026). “—” indicates the distribution was not reported in that wave.

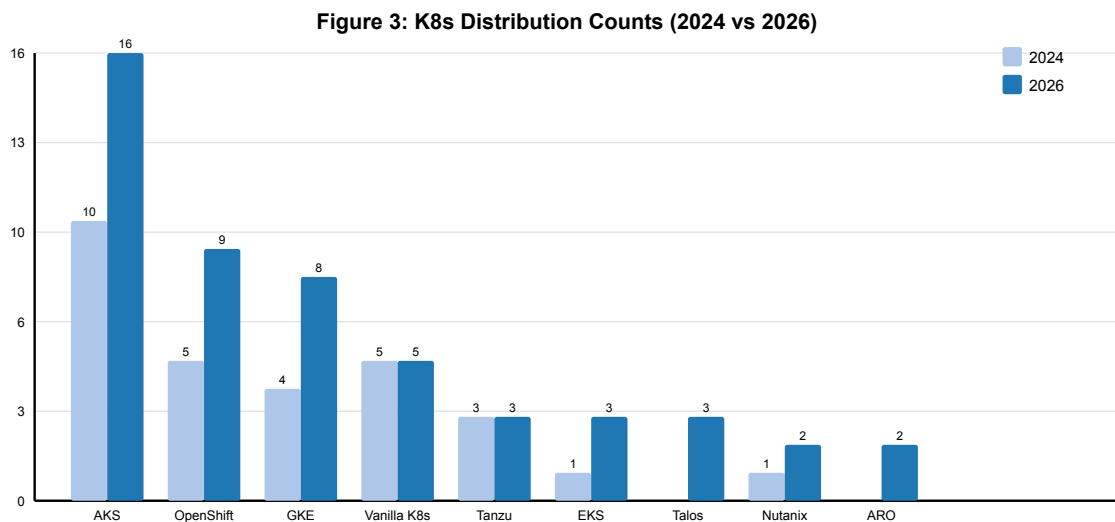


Figure 3: Kubernetes Distribution Adoption Counts (2024 vs 2026)

- **Infrastructure and Kubernetes Distributions:** Organizations are actively offloading control-plane management. While AKS grew in absolute terms (10 to 16 entries), its relative share declined slightly from 45% to 40% as the Kubernetes-using population nearly doubled (n=22 to n=40). Vanilla Kubernetes installations remained flat at 5 but declined sharply in share from 23% to 12% (-11pp), confirming the managed-service shift. Packaged enterprise distributions like OpenShift maintained their share (23%→22%), while new entrants in 2026 include Talos (an immutable, API-driven OS for Kubernetes, 8%) and ARO (Azure Red Hat OpenShift, 5%).
- **Cloud Providers:** Among platform entries naming a primary cloud provider (n=24 in 2024, n=34 in 2026), Azure remains the dominant primary cloud provider (59%, 20/34 in 2026, down slightly from 63%, 15/24), followed by Google Cloud (24%, 8/34, up from 21%), and AWS (18%, 6/34, up from 17%).
- **CI/CD and Deployment (Table 5):** The deployment toolchain shows clear consolidation around GitOps. GitHub Actions rose from 32% to 52% (+20pp) and ArgoCD from 26% to 50% (+24pp), with

both crossing the majority threshold. Jenkins declined sharply from 24% to 8% (−16pp), and Flux declined from 24% to 15% (−9pp).

Table 5: Deployment and CI/CD Tool Adoption

Tool	2024 (n=34)	2026 (n=48)	Δ (pp)
GitHub Actions	32% (11/34)	52% (25/48)	+20
ArgoCD	26% (9/34)	50% (24/48)	+24
Azure DevOps	18% (6/34)	17% (8/48)	−1
GitLab Pipelines	15% (5/34)	15% (7/48)	0
Flux	24% (8/34)	15% (7/48)	−9
kubectl apply	12% (4/34)	15% (7/48)	+3
Jenkins	24% (8/34)	8% (4/48)	−16

Multi-select: platform entries may report multiple tools. Base is entries reporting at least one deployment tool (n=34 in 2024, n=48 in 2026). The growth of ArgoCD (26%→50%) alongside GitHub Actions (32%→52%) reflects a separation of CI (GitHub Actions) and CD (ArgoCD) concerns.

- **Infrastructure as Code (Table 6):** Terraform strengthened its dominant position, rising from 73% to 82% (+9pp). Azure Bicep grew from 6% to 16% (+10pp) while ARM Templates declined from 15% to 7% (−8pp), indicating an Azure-native IaC migration pattern. Crossplane (Kubernetes-native IaC) grew from 3% to 9%, though the small absolute base (1→4 entries) warrants caution.

Table 6: Infrastructure as Code Tool Adoption

Tool	2024 (n=33)	2026 (n=45)	Δ (pp)
Terraform	73% (24/33)	82% (37/45)	+9
Ansible	42% (14/33)	38% (17/45)	−4
Azure Bicep	6% (2/33)	16% (7/45)	+10
Pulumi	6% (2/33)	9% (4/45)	+3
Crossplane	3% (1/33)	9% (4/45)	+6
ARM Templates	15% (5/33)	7% (3/45)	−8
AWS CloudFormation	9% (3/33)	4% (2/45)	−5

Multi-select: platform entries may report multiple tools. Base is entries reporting at least one IaC tool (n=33 in 2024, n=45 in 2026).

- **Observability and Monitoring:** The observability stack is maturing. While Grafana (28→38) and Prometheus (22→35) remain foundational, the strongest growth is in OpenTelemetry (9→22 mentions), suggesting a shift toward vendor-neutral, standards-based telemetry. The Grafana stack saw broad adoption: Loki (6→21 mentions), Tempo (4→14), and Mimir (5→8) indicate organizations are building integrated observability stacks rather than relying on point solutions.
- **Security Tooling:** Consistent with the rise of security as a primary motivation (Section 4.2), the security tooling mix evolved. For network security, Cilium grew from 3 to 13 mentions, overtaking Linkerd (4→2) as the most widely adopted CNI/service mesh. For application security and policy enforcement, Kyverno grew from 4 to 13 mentions, becoming the leading policy engine ahead of OPA (4→6). Trivy (0→4) also emerged as a new container scanning tool. Tool mentions are counted

per question and by exact match: Cilium is named under network security by 3 and 13 entries; one 2026 entry also writes “CiliumNetworkPolicies”, which the exact-match rule does not count.

- **Kubernetes Operational Practices (2026 only):** The 2026 wave included new questions on Kubernetes operational depth. Of the 40 K8s-using entries: 53% (21/40) develop custom Kubernetes operators, indicating deep API-level integration rather than surface-level adoption. For databases in Kubernetes, 58% (23/40) do not run databases on Kubernetes, while 13 (33%) do. Among these, CloudNativePG is named by 6 entries and the Zalando Postgres Operator by 5; the remaining 4 answers describe proof-of-concept work or are unclear (see Section 3.3 for coding).
- **Self-Service Capabilities:** The breadth of platform self-service offerings expanded. While build, deploy, and observability capabilities (logs, metrics, dashboards) were already common in 2024 (87–100% of platforms), the largest growth occurred in databases and messaging/storage (47%, 14/30 → 86%, 38/44; $p < 0.001$, Fisher’s exact test) and security/vulnerability scanning (60%, 18/30 → 75%, 33/44; $p = 0.206$). The database self-service increase is the largest change in the dataset and the only one below every Bonferroni threshold that could be applied to the comparisons in this paper. One capability declined: build and deploy fell from 100% (30/30) to 91% (40/44; $p = 0.142$). This broadening aligns with the shift toward platforms providing data and security services alongside core compute.
- **Programming Language Diversity:** Analysis of the top three programming languages per platform indicates a diversifying workload. While enterprise mainstays Java (23% to 19% of mentions) and Kotlin (19% to 17%) experienced slight relative declines, TypeScript (13% to 17%) and Python (14% to 17%) rose to parity with Kotlin. Rust also made its first appearance in the 2026 dataset (2%).

5. Discussion

The data from 2024 to 2026 suggest that platform engineering among surveyed organizations has moved past initial adoption and into the operational demands of running enterprise-scale infrastructure.

5.1. From Innovation to Standard Utility

In 2024, platform engineering among the surveyed organizations was associated with agility and, to a lesser extent, attracting engineering talent. By 2026, the data suggest that platforms are increasingly viewed as operational infrastructure rather than modernization projects. The 25pp increase in security as a primary motivation—the only motivation change reaching $\alpha = 0.05$ at the individual test level (full-sample Fisher $p = 0.034$, OR=3.4 for 2026 relative to 2024; paired panel McNemar $p = 0.021$; see Section 6 for multiple comparisons)—together with directional increases in cost optimization (+15pp) and quality (+9pp), is consistent with platforms maturing into infrastructure expected to enforce governance and manage resources. This interpretation is corroborated by the tooling data: the rapid growth of Cilium (3→13 mentions) and Kyverno (4→13) in the security tooling data (Section 4.4) suggests that organizations are not merely stating security as a priority but actively investing in enforcement infrastructure. An external regulatory factor likely reinforces this trend: the EU NIS2 Directive (2022/2555) [13], which entered into force in January 2023 and whose transposition deadline for EU member states passed in October 2024, imposes stricter requirements for cybersecurity risk management, supply chain security, and incident reporting on essential and important entities including public administration bodies. Although NIS2 has not yet been transposed into Norwegian law — Norway implemented the predecessor NIS1 directive through the Digital Security Act (digitalsikkerhetsloven), which entered into force on 1 October 2025 [29] — Norwegian public sector organizations were actively preparing for NIS2’s anticipated requirements during the study period, as the directive must eventually be incorporated via the EEA Agreement. The timing of the security motivation increase (from 57% in the 2024 baseline, collected before the NIS2 transposition deadline, to 82% in the 2026 wave) aligns with anticipatory regulatory pressure amplifying pre-existing organizational interest in platform-level security enforcement. However, we note that this interpretation rests on temporal coincidence and self-

reported motivations, not observed compliance outcomes; the survey measures what organizations say drives their platform investment, not the actual governance or security effects achieved.

Viewed through the lens of Rogers’ Diffusion of Innovations [9], the 86%→92% IDP adoption growth places this community in the late majority stage: the remaining non-adopters likely face structural barriers (regulatory constraints, legacy systems) rather than awareness gaps. Accordingly, the discourse within this community has shifted from *whether* to adopt platforms to *what those platforms should prioritize* — a pattern consistent with Rogers’ characterization of late-stage adoption as driven by institutional pressure and risk reduction rather than competitive advantage.

The tooling convergence observed in RQ4 can be analyzed more precisely through DiMaggio and Powell’s [10] three mechanisms of institutional isomorphism. Mimetic isomorphism is evident in the technology diffusion pattern: early-adopting agencies such as Nav and Skatteetaten, which are among the largest and most visible platform operators in the community, serve as organizational templates whose publicly documented platforms [17][18] and practices propagate through the Offentlig PAAS community’s regular knowledge-sharing sessions. Normative isomorphism operates through professional networks — as platform engineers move between agencies, they carry tooling preferences and operational practices. Coercive isomorphism may operate through shared procurement frameworks, such as the state’s shared cloud procurement arrangements (Markedsplassen for skytjenester) [30], which may contribute to the concentration on a few large providers, including Azure’s sustained dominance (59% of entries naming a primary cloud provider); we did not examine individual procurement decisions. The convergence on a narrow toolset — rather than the diversity one might expect from independent technology evaluations across 45 organizations — suggests that institutional isomorphism plays a significant role in technology selection alongside (or instead of) purely rational evaluation.

5.2. The Measurement Gap

The data show a divergence between operational and measurement maturity. While the modal operational maturity shifted from Level 2 (Centrally Tracked) to Level 3 (Centrally Enabled)—a transition that coincides temporally with widespread adoption of declarative GitOps tooling—the ability to measure platform impact remains largely ad hoc. Nearly half of all platform entries (49%, 21/43) report Level 1 for Measurement in 2026.

This gap grows more pressing as organizations shift their platform mandates toward cost optimization (as the motivation data suggest): the inability to quantitatively measure platform efficiency could create friction between engineering teams and organizational leadership. We note that the survey measures self-assessed maturity levels for measurement practices but does not directly assess whether organizations employ structured measurement frameworks (such as GQM [5]). The persistence of Level 1 measurement may reflect multiple factors: the inherent difficulty of defining platform metrics, competing priorities during operational scaling, or the absence of organizational incentives to quantify platform value when funding is already secured.

This tension can be partially understood through the lens of principal-agent theory. Platform teams (agents) possess specialized knowledge about platform operations that organizational leadership (principals) cannot directly observe. In the absence of mature measurement practices, this information asymmetry favors platform teams—they can assert value without quantitative evidence. Developing mature measurement practices (Level 3+) could expose underperformance or reveal that platform return on investment is lower than assumed. This creates a potential misalignment of incentives: the teams best positioned to develop measurement capabilities have limited incentive to do so, particularly when platform investment is already secured (all 43 entries report Investment at Level 2 or above, indicating at least dedicated cost centers).

We also note that measurement maturity may be an inherently trailing indicator—organizations may need to achieve operational stability before investing in sophisticated metrics practices. The emergence

of a small vanguard at Level 3 (12%, 5/43 in 2026, up from 0% in 2024) may signal the beginning of this transition, though the absolute numbers are too small to draw firm conclusions.

5.3. Blind Spots: The Missing Voice of the Developer

While the data provides a view of how platform *builders* (Platform Developers, DevOps, Architects) rate their initiatives, it lacks the perspective of the platform *users* (the application developers). As noted in the original 2024 baseline study, a qualitative blind spot is user sentiment. The SPACE framework [14] identifies satisfaction and well-being as a key dimension of developer productivity that cannot be captured through organizational-level surveys of platform teams. Similarly, recent work on developer experience [15] emphasizes that productivity gains are experienced at the individual level, not the organizational level at which this study operates. Without correlating the builders' self-assessed maturity against actual developer satisfaction (e.g., via developer experience surveys or cognitive load assessments following the SPACE dimensions), it is not possible to confirm that these platforms are achieving their primary stated goal of reducing friction for end-users. This gap matters: Adoption maturity at Level 4 (Participatory contribution) declined from 13% to 2%, suggesting that while platforms may be operationally mature, the developer community may not yet be deeply engaged in platform evolution.

6. Threats to Validity and Limitations

We structure this section following the validity categories proposed by Petersen and Gencel [19] for empirical software engineering research, supplemented by the survey-specific guidance of Molléri et al. [22] and the ACM SIGSOFT Empirical Standards for questionnaire surveys [23].

6.1. Construct Validity

- **Maturity Model Operationalization:** The CNCF Platform Maturity Model [4] originates from practitioner consensus rather than theoretically grounded scholarly work. Its conceptual validity and underlying assumptions have not been independently validated in peer-reviewed research. We adopted it because (a) no alternative validated instrument exists for platform engineering maturity, (b) it is the de facto community standard, enabling practical relevance, and (c) its five-aspect structure provides multidimensional assessment rather than a single aggregate score. However, respondents' interpretations of maturity levels may vary, particularly for the "Operations" aspect, which the 2024 baseline study noted was perceived as complex. The increase in Level 1 for Operations (from 13% to 19%), including three returning organizations that moved from Level 2 to Level 1, may reflect interpretation shifts rather than actual regression.
- **Self-Reporting Bias:** All maturity assessments, platform scale, and motivation data are self-reported. We mitigated inflation by normalizing responses to favor the most technical, hands-on roles within an organization, but independent calibration of self-assessed levels was not performed. Self-reported organizational scale (team counts) and motivations were not validated against external data.

6.2. Internal Validity

- **Cohort Composition Effects:** This study employs a repeated cross-sectional design: 27 of the 45 organizations in the 2026 wave (60%) also participated in 2024, while 18 are new entrants and 8 from 2024 did not return. This design cannot fully distinguish temporal change from cohort composition effects. However, a sensitivity analysis of the 27-organization panel subset shows directional trends consistent with the full-sample results across all key findings. Using paired McNemar tests on the 24 organizations with a platform in both waves: security motivation (54%→88%, 13→21, $p=0.021$), Operations L3+ (38%→54%, 9→13, $p=0.344$), cost motivation (42%→54%, 10→13, $p=0.453$), and Measurement L1 persistence (42%→50%, 10→12, $p=0.688$). None of these panel-only comparisons diverge directionally from the full-sample trends, lending confidence that the observed shifts are not artifacts of cohort composition. Additionally, a comparison of new entrants against returning

organizations shows no systematic bias: new entrants report lower platform adoption (84% vs 97%) and comparable maturity distributions.

- **Independence of Observations:** Statistical tests assume independent observations. Three organizations in the 2026 wave operate dual platforms, yielding 6 of 48 entries (12.5%) that may share governance, funding, or strategic priorities. For organizational-level constructs (motivations, investment), these paired entries are not fully independent. We assess this as unlikely to materially affect results given the small proportion, but note that a sensitivity analysis excluding dual-platform entries produces directionally identical results for all key findings.
- **Researcher Positionality:** This is insider research [31]: the author is both a practitioner in and the chair of the community studied. The author designed the survey, distributed it through the Offentlig PAAS community (which he chairs), and performed all data collection, normalization, and analysis. While the normalization rules were defined a priori (Section 3.3), the single-researcher design means that analytical choices were not independently validated. To mitigate this, we publish the analysis script and aggregate frequency tables (see Data Availability) and have documented all merge and coding decisions (Section 3.3).

6.3. Conclusion Validity

- **Statistical Power:** The sample sizes ($n=35$ for 2024, $n=48$ for 2026) limit the statistical power of cross-wave comparisons. Most observed percentage-point changes do not reach statistical significance at $\alpha=0.05$ using Fisher's exact test. Among the seven motivations, only the security increase ($p=0.034$) is individually significant, and it would not survive correction for multiple comparisons (Bonferroni-corrected $\alpha=0.007$ for 7 simultaneous tests). The database and messaging self-service increase ($p<0.001$) is the only change that survives such a correction. We therefore present most cross-wave changes as observed directional trends rather than confirmed effects, and caution against over-interpreting individual shifts. The inferential statistics should be treated as exploratory rather than confirmatory.
- **Overlapping Samples:** About 60% of the organizations appear in both waves, so the full-sample Fisher tests compare partly overlapping rather than independent samples. Positive correlation within returning organizations generally makes such tests conservative, but the reported p-values are approximate. For this reason the panel comparisons use paired McNemar tests (Section 6.2).

6.4. External Validity

- **Sampling and Selection Bias:** The survey was distributed via the *Offentlig PAAS* community network (Slack and LinkedIn). This methodology inherently targets organizations that are already actively engaged in platform engineering discourse. Consequently, the 92% adoption rate reflects penetration within this digitally engaged community, not a randomized sample of all Norwegian public-sector organizations. Smaller agencies that primarily rely on outside vendors for operations are likely underrepresented. The community had 84 member organizations at the time of the 2026 wave, of which 45 responded (54% organizational response rate). Organizations joining and leaving the community between the waves were not tracked, so we cannot assess how much the sampling frame changed.
- **Generalizability:** The Norwegian public sector is uniquely characterized by high funding, strong labor unions, and a mandate for digital public services [24]. The results reflect the experiences of *surveyed organizations within the Offentlig PAAS community* and should not be generalized uncritically to all Norwegian public-sector organizations, nor to private-sector enterprises or other national contexts. While the technological trends (Kubernetes, GitOps) are broadly applicable given their global adoption patterns, the organizational dynamics (motivations, maturity trajectories) may be specific to this well-funded, early-adopting population.

7. Conclusion and Future Work

Over the past two years, Internal Developer Platforms have become the dominant operating model for software delivery in the surveyed Norwegian public sector community. The data suggest a directional shift from agility and talent recruitment toward security enforcement and cost optimization, though among the motivations only the security increase reaches $\alpha=0.05$ at the individual test level (uncorrected; paired panel $p=0.021$). The technical toolchain has converged on a standardized stack centered on managed Kubernetes, declarative infrastructure (Terraform), and GitOps (ArgoCD, GitHub Actions).

The most actionable finding is the persistent gap between operational and measurement maturity. While organizations report advancing their operational practices, nearly half remain at Level 1 for measurement. As platforms are increasingly expected to justify their cost and security contributions, closing this measurement gap is the clearest next step for the community.

Future Work: Future research should prioritize three directions. First, qualitative studies — such as developer interviews and cognitive load assessments following the SPACE framework [14] — are needed to bridge the gap between how platforms are built and how they are experienced by their users. Second, independent calibration of self-assessed maturity levels (e.g., through organizational case studies) would strengthen the validity of maturity model research. Third, a dedicated panel study tracking the same organizations across three or more waves would enable true within-subject longitudinal analysis, avoiding the cohort composition limitations of the present study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The author conducted this study as part of his employment at the Norwegian Labour and Welfare Administration (Nav).

Acknowledgements

The author thanks Audun Fauchald Strand and Are Vattekar, who founded the Offentlig PAAS community of practice [16] after meeting at KubeCon Europe in Berlin in 2017, and whose early platform engineering work at Nav (NAIS [17]) and Skatteetaten (Aurora [18]) starting in 2016 helped catalyze adoption across the Norwegian public sector. The cross-agency network they built — about 1,800 members at the time of the 2026 wave — provided the sampling infrastructure and professional relationships that made this longitudinal study feasible. The author also thanks every survey respondent in both waves and the organizations that permitted their participation.

CRediT Author Statement

Hans Kristian Flaatten: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Software, Visualization, Writing – original draft, Writing – review and editing, Project administration.

Declaration of Competing Interests

The author is employed by Nav, one of the organizations included in the survey, and serves as chairman of Offentlig PAAS, the not-for-profit community of practice through which the survey was distributed. Both the survey design and the normalization rules were defined independently of these roles.

Data Availability

Respondents were told that aggregated results would be published (Section 3.4). We therefore publish aggregate frequency tables per question for each wave, not response-level data. Cells with fewer than three platform entries are shown as “<3”. The released tables cross-tabulate many attributes per wave,

where small cells could combine into an identifying profile; the paper reports single-attribute marginals only, and there we give exact counts of 1–2 because they are needed to interpret the analysis and do not identify an organization on their own. The tables are available at <https://github.com/offentlig-paas/open-data/tree/main/state-of-platforms/2024> and <https://github.com/offentlig-paas/open-data/tree/main/state-of-platforms/2026>. The analysis script is available at <https://github.com/offentlig-paas/open-data/tree/main/state-of-platforms/scripts>; it requires the response-level data, which is not published. The survey instrument for each wave is at <https://github.com/offentlig-paas/open-data/tree/main/state-of-platforms/2024/instrument.md> and <https://github.com/offentlig-paas/open-data/tree/main/state-of-platforms/2026/instrument.md>.

Declaration of Generative AI and AI-Assisted Technologies in the Manuscript Preparation Process

During the preparation of this work the author used GitHub Copilot (Claude Opus 4.6) in order to assist with editing, statistical verification, and prose revision. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article. The tool was not used for data collection or analysis.

References

- [1] M. Skelton and M. Pais, *Team Topologies: Organizing Business and Technology Teams for Fast Flow*. Portland, OR: IT Revolution Press, 2019.
- [2] N. Forsgren, J. Humble, and G. Kim, *Accelerate: The Science of Lean Software and DevOps: Building and Scaling High Performing Technology Organizations*. Portland, OR: IT Revolution Press, 2018.
- [3] DORA Team, “State of DevOps Report,” Google Cloud, 2024. [Online]. Available: <https://dora.dev/research/> (accessed Mar. 17, 2026).
- [4] CNCF TAG App Delivery, “Platform Engineering Maturity Model,” Cloud Native Computing Foundation, 2024. [Online]. Available: <https://tag-app-delivery.cncf.io/whitepapers/platform-eng-maturity-model/> (accessed Mar. 17, 2026).
- [5] V. R. Basili, G. Caldiera, and H. D. Rombach, “The Goal Question Metric Approach,” in *Encyclopedia of Software Engineering*, J. J. Marciniak, Ed. New York: Wiley, 1994, pp. 528–532.
- [6] Humanitec, “State of Platform Engineering Report, Vol. 3,” 2024. [Online]. Available: <https://humanitec.com/whitepapers/state-of-platform-engineering-report-volume-3> (accessed Mar. 17, 2026).
- [7] Puppet, “2023 State of DevOps Report: Platform Engineering Edition,” Perforce Software, 2023. [Online]. Available: <https://www.puppet.com/resources/state-of-devops-report> (accessed Mar. 17, 2026).
- [8] CNCF End User Technology Radar, “Platforms,” Cloud Native Computing Foundation, 2024. [Online]. Available: https://www.cncf.io/reports/?_sft_lf-report-type=radar (accessed Oct. 12, 2026).
- [9] E. M. Rogers, *Diffusion of Innovations*, 5th ed. New York: Free Press, 2003.
- [10] P. J. DiMaggio and W. W. Powell, “The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields,” *American Sociological Review*, vol. 48, no. 2, pp. 147–160, 1983.
- [11] J. S. Molléri, K. Petersen, and E. Mendes, “Survey Guidelines in Software Engineering: An Annotated Review,” in *Proc. 10th ACM/IEEE Int. Symp. on Empirical Software Engineering and Measurement (ESEM)*, Ciudad Real, Spain, 2016, pp. 1–6.
- [12] B. Kitchenham and S. L. Pfleeger, “Personal Opinion Surveys,” in *Guide to Advanced Empirical Software Engineering*, F. Shull, J. Singer, and D. I. K. Sjøberg, Eds. London: Springer, 2008, pp. 63–92.
- [13] European Parliament and Council of the European Union, “Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2),” *Official Journal of the European*

- Union*, L 333, pp. 80–152, Dec. 2022. [Online]. Available: <https://eur-lex.europa.eu/eli/dir/2022/2555> (accessed Mar. 18, 2026).
- [14] N. Forsgren, M.-A. Storey, C. Maddila, T. Zimmermann, B. Houck, and J. Butler, “The SPACE of Developer Productivity: There’s More to It Than You Think,” *ACM Queue*, vol. 19, no. 1, pp. 20–48, 2021.
- [15] A. Noda, M.-A. Storey, N. Forsgren, and M. Greiler, “DevEx: What Actually Drives Productivity,” *ACM Queue*, vol. 21, no. 2, pp. 35–53, 2023.
- [16] Offentlig PAAS, “Om Offentlig PAAS” [About Offentlig PAAS], 2024. [Online]. Available: <https://offentlig-paas.no/om> (accessed Mar. 18, 2026).
- [17] Nav, “NAIS — NAVs Application Infrastructure Services,” 2024. [Online]. Available: <https://nais.io> (accessed Mar. 18, 2026).
- [18] Skatteetaten, “Aurora,” 2024. [Online]. Available: <https://skatteetaten.github.io/aurora/> (accessed Mar. 18, 2026).
- [19] K. Petersen and C. Gencel, “Worldviews, Research Methods, and their Relationship to Validity in Empirical Software Engineering Research,” in *Proc. Joint Conf. of the 23rd Int. Workshop on Software Measurement and 8th Int. Conf. on Software Process and Product Measurement (IWSM-Mensura)*, Ankara, Turkey, 2013, pp. 81–89.
- [20] K.-J. Stol and B. Fitzgerald, “The ABC of Software Engineering Research,” *ACM Transactions on Software Engineering and Methodology*, vol. 27, no. 3, Article 11, 2018.
- [21] L. Leite, C. Rocha, F. Kon, D. Milojicic, and P. Meirelles, “A Survey of DevOps Concepts and Challenges,” *ACM Computing Surveys*, vol. 52, no. 6, Article 127, pp. 1–35, 2019.
- [22] J. S. Molléri, K. Petersen, and E. Mendes, “An Empirically Evaluated Checklist for Surveys in Software Engineering,” *Information and Software Technology*, vol. 119, Article 106240, 2020.
- [23] P. Ralph et al., “Empirical Standards for Software Engineering Research,” arXiv:2010.03525, 2021. [Online]. Available: <https://www2.sigsoft.org/EmpiricalStandards/> (accessed May 10, 2026).
- [24] OECD, “Digital Government Index: 2019 Results,” *OECD Public Governance Policy Papers*, No. 03, OECD Publishing, Paris, 2020.
- [25] CNCF TAG App Delivery, “CNCf Platforms White Paper,” Cloud Native Computing Foundation, 2023. [Online]. Available: <https://tag-app-delivery.cncf.io/whitepapers/platforms/> (accessed May 10, 2026).
- [26] M. Janssen and A. Joha, “Challenges for Adopting Cloud-Based Software as a Service (SaaS) in the Public Sector,” in *Proc. European Conf. on Information Systems (ECIS)*, 2011.
- [27] Q. McNemar, “Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages,” *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.
- [28] The National Committee for Research Ethics in the Social Sciences and the Humanities (NESH), *Guidelines for Research Ethics in the Social Sciences and the Humanities*, 5th ed. Oslo: Forskningsetikkomiteene, 2021. [Online]. Available: <https://www.forskningsetikk.no/en/guidelines/social-sciences-and-humanities/> (accessed Oct. 12, 2026).
- [29] Lovdata, “Lov om digital sikkerhet (digitalsikkerhetsloven),” LOV-2023-12-20-108. [Online]. Available: <https://lovdata.no/dokument/NL/lov/2023-12-20-108> (accessed Oct. 12, 2026).
- [30] DFØ, “Markedsplassen for skytjenester” [Marketplace for cloud services]. [Online]. Available: <https://markedsplassen.anskaffelser.no/> (accessed Oct. 12, 2026).
- [31] J. Singer and N. G. Vinson, “Ethical Issues in Empirical Studies of Software Engineering,” *IEEE Transactions on Software Engineering*, vol. 28, no. 12, pp. 1171–1180, 2002.