DYRÍGO

THINKLAB

BEYOND THE SCORE

OECD Digital Government Index — Benchmark Integrity Review

Capability • Governance • Ranking • Human Delivery • Outcome • Cost

FINAL VERDICT

The OECD Digital Government Index should remain available as a benchmark of institutional digital-government capability. The public comparative User-Driven ranking, however, should not continue in its current form as an unqualified country ranking. Its label and downstream use can reasonably be read as evidence of realised user-centred performance, while the measurement and validation architecture does not establish that meaning. OECD should suspend or clearly withdraw the current User-Driven country ranking from performance-style presentation until it is either independently validated against realised human outcomes or redesigned and relabelled so that its public meaning cannot exceed its evidentiary basis.

RESEARCH CLOSED • EVIDENCE RECONCILED • EXTERNAL REPORT

Contents

Executive summary....3
1. Objective, scope and decision standard....5
2. The benchmark OECD actually built....6
3. From government evidence to public ranking....8
4. Findings: where capability stops and human reality begins....10
5. The decisive cross-country test....13
6. What a national score can make invisible....16
7. Human reality: an end-to-end mechanism case....17
8. The value of a high score....19
9. When a capability score acquires a larger meaning....21
10. Governance and independence....22
11. Mission, purpose and the human beneficiary....23
12. What the DGI is good for — and what it should not be used for....24
13. Publication decision: what should happen to the list now?....25
14. Recommended redesign....27
15. Counter-evidence, boundaries and what this report does not claim....29
16. Final conclusion and advice to OECD....30
17. Primary sources....32
Appendix A. Material evidence map....34
Appendix B. Publication language guardrails....36

Executive summary

A benchmark can be internally robust and still be externally over-interpreted.

This review examined the validity, reliability, transparency and practical meaning of the OECD Digital Government Index (DGI), with special attention to the User-Driven dimension. The question was not whether digital-government capability matters. It does. The question was whether the public meaning of the score remains within what the evidence can actually support when the score travels from institutional evidence to ranking, political communication and claims about people's lived experience.

RESEARCH OBJECTIVE

Determine how far the DGI can legitimately be interpreted along the chain CAPABILITY → GOVERNANCE → SCORE → SERVICE → END-TO-END JOURNEY → REALISED HUMAN OUTCOME → COST / LOST VALUE.

The review finds that OECD has built a credible and useful capability benchmark. OECD itself now describes the DGI as an assessment of enabling foundations for coherent digital transformation rather than a measure of the digitalisation of specific services. That clarification is decisive. It defines what the index can validly claim — and where its evidentiary reach ends.

The User-Driven dimension is not empty, cosmetic or merely documentary. It captures real governmental practices: user engagement, service standards, testing, feedback, implementation and monitoring. These capabilities can contribute to better services. But contribution is not equivalence. A government can possess mature user-centred capabilities while citizens still experience fragmented access, repeated proof requirements, handoff failures, recovery burden and high time-to-outcome.

OECD's own 2026 Digital Government Outlook confirms this gap. It states that the challenge is now to translate foundations into tangible impact, that service standards are inconsistently applied, that joined-up delivery remains incomplete, and that only 28% of OECD countries report standardised mechanisms for measuring the transaction costs services impose on users — including time, steps, documents, handovers and repeat contact after failure.

The OECD Trust Survey provides independent evidence about experienced administrative interactions. It is not a substitute measure for DGI User-Driven and does not measure the same construct. Its value here is as an external contradiction test: if User-Driven is allowed to function publicly as a proxy for realised user-centred performance, country ordering should have a defensible relationship with independent user evidence. Portugal scores 0.94 on User-Driven but reports 52% satisfaction among recent users of administrative services; Korea scores 0.91 and reports 79%; Estonia scores materially lower on User-Driven while its administrative-service satisfaction is above 80%; Chile combines a very high User-Driven score with markedly weaker

satisfaction. These inversions do not prove that User-Driven is wrong. They show that an outcome-performance meaning cannot be assumed without external validation.

A second external test reaches the same boundary from a different measurement layer. The European Commission's 2025 eGovernment Benchmark observes the online delivery of major public services much more directly than the DGI capability architecture. On the Commission's 'digital public services for citizens' measure, Estonia scores 96/100, the Netherlands 89/100, Spain 89/100 and Portugal 84/100. The corresponding 2025 OECD User-Driven ordering is Portugal 0.94, Spain 0.73, Estonia 0.72 and the Netherlands 0.56. The capability ranking therefore does not reproduce the ordering on this substantially more service-observational European measure. The EU benchmark is not itself a complete human-outcome measure; it strengthens the need to keep capability, observed service delivery and realised human outcome analytically separate.

The review also identifies a second integrity layer. High rankings have reputational, political and institutional value. Governments can present them as evidence of leadership, policy success and international standing; rankings can become strategic objectives and can influence investment priorities. Economic value is plausible through signalling, attractiveness and policy legitimacy, but this review does not claim a quantified causal economic return where such causality has not been established. The material point is that a valuable score creates incentives to optimise what is observable and rewarded. That makes independent outcome validation more important, not less.

THE CENTRAL INTEGRITY PROBLEM

EVIDENCE–CLAIM DISTANCE: a capability score can remain technically valid while acquiring, through its label, ranking, authority and downstream use, an outcome meaning that the underlying measurement has not established.

The report therefore recommends a differentiated remedy. The DGI composite should not be discarded; it is useful as a capability benchmark. The current standalone/performance-like presentation of the User-Driven country ranking, however, should not continue in its present form. OECD should retain the capability measure, relabel its construct explicitly as capability, and pair comparative use with independent service-delivery and human-outcome evidence. Formal suspension is one possible transition mechanism, but the evidence establishes the need to change the current presentation rather than prescribing a single administrative mechanism.

WHAT SHOULD HAPPEN NOW?
KEEP THE DGI CAPABILITY BENCHMARK. DISCONTINUE THE CURRENT STANDALONE/PERFORMANCE-LIKE USER-DRIVEN COUNTRY-RANKING PRESENTATION IN ITS PRESENT FORM. RELABEL THE CONSTRUCT EXPLICITLY AS CAPABILITY AND PAIR COMPARATIVE USE WITH INDEPENDENT SERVICE-DELIVERY AND HUMAN-OUTCOME EVIDENCE.

1. Objective, scope and decision standard

The research objective governs the conclusion — not the other way around.

1.1 Frozen objective

Systematically assess the validity, reliability, transparency and practical meaning of the OECD Digital Government Index and related benchmarking claims, with the central chain capability → governance → score → end-to-end service → realised user outcome. Establish to what extent the DGI measures genuinely user-centred public-service performance versus institutional and policy capability.

1.2 Reality chain

REALITY CHAIN

INFRASTRUCTURE → CAPABILITY → GOVERNANCE → SCORE → SERVICE → ACCESS → JOURNEY → FAILURE / RECOVERY → REALISED HUMAN OUTCOME → COST / LOST VALUE

1.3 Questions the review must answer

- What does the DGI actually measure, and what does it explicitly not measure?
- How is evidence supplied, validated, scored, weighted and converted into a country ranking?
- Does User-Driven measure observed user outcomes or the institutional conditions intended to produce them?
- Do independent outcome measures reproduce the country ordering implied by User-Driven?
- What human burden remains invisible when measurement stops at capability, portal or transaction level?
- Who benefits from a high score, what political, institutional, reputational or economic value can it create, and what incentives follow?
- How do OECD labels and rankings travel through national communication and downstream policy use?
- How independent is the evidence and validation architecture from the entities being ranked?
- What should OECD do when the public meaning of a ranking exceeds what its evidence validates?

1.4 Decision standard

The remedy is proportional to the demonstrated integrity problem. Withdrawal is not justified merely because a benchmark is imperfect. It becomes justified where continued publication in its current form creates a foreseeable and material interpretation that the evidence does not validate, and where a lighter intervention would not adequately restore the boundary between measurement and claim.

2. The benchmark OECD actually built

The strongest critique starts by describing the benchmark accurately.

The 2025 DGI assesses enabling foundations for digital transformation across six dimensions: Digital by Design, Data-driven Public Sector, Government as a Platform, Open by Default, User-Driven and Proactiveness. OECD's 2026 country notes explicitly state that the DGI does not measure the digitalisation of specific services; it focuses on strategies, policy levers, implementation practices and monitoring mechanisms that enable coherent whole-of-government transformation.

2.1 This is a strength, not a defect

A government needs architecture, governance, data, platforms, standards, user-engagement mechanisms and monitoring capability before it can reliably deliver good services at scale. Measuring those foundations is legitimate. The review therefore rejects any claim that the DGI is worthless simply because it is not an outcome index.

2.2 The boundary becomes critical at User-Driven

The term User-Driven is semantically powerful. In technical context it describes governmental capability and practice. In ordinary language it can be heard as a judgement about whether government actually works for users. That difference is the beginning of the integrity problem: the label is broader than the measurement layer that supports it.

CONSTRUCT BOUNDARY

USER-CENTRED CAPABILITY ≠ OBSERVED USER-CENTRED PERFORMANCE

2.3 2023 to 2025: replication, not disappearance

The 2025 edition does not remove the core issue identified in the earlier edition. OECD's later language actually clarifies it: the DGI measures foundations, while the 2026 Outlook separately focuses on the challenge of translating those foundations into tangible impact. The distinction is therefore temporal as well as conceptual: the benchmark has matured, but the capability-to-outcome boundary remains.

2.4 Selected 2025 scores

Country	DGI composite	User-Driven
Korea	0.95	0.91
Australia	0.88	0.95

Portugal	0.86	0.94
Estonia	0.83	0.72
Chile	0.79	0.90
Spain	0.75	0.73
Netherlands	0.65	0.56

3. From government evidence to public ranking

Validation can be rigorous within a construct without validating a different construct.

The DGI is built through a substantial evidence-validation architecture. In the 2023 methodology, 94 questions were addressed primarily to high-level digital-government officials. OECD describes systematic matching of responses and supporting evidence to questionnaire requirements, followed by a second transversal-consistency review. Where answers could not be validated, the Secretariat could request clarification or further evidence and could validate or amend responses with a documented rationale; participating countries formally approved their final responses. The 2025 Data Explorer further states that supporting documentation was systematically required, that the OECD Secretariat conducted thorough validation and fact-checking, and that responses without supporting evidence were not considered. OECD also applies statistical validation, including correlation analysis, principal-component analysis, internal-reliability testing and sensitivity analysis. These are meaningful and comparatively strong controls.

MEASUREMENT PIPELINE

GOVERNMENT INPUT → OECD EVIDENCE VALIDATION → DATA CONFIRMATION → SCORING → WEIGHTING → COUNTRY SCORE → DIMENSION LABEL → RANKING → PUBLIC INTERPRETATION

3.1 What validation proves

- That submitted evidence meets the benchmark's definitions and scoring requirements.
- That the index is internally coherent enough to support comparison within its intended construct.
- That the capability construct and its scoring architecture can be tested for internal reliability, dimensional structure and sensitivity to methodological choices.

3.2 What it does not automatically prove

- That a citizen can complete an end-to-end journey successfully.
- That services are equally accessible across geography, channels or user groups.
- That failure and recovery paths work.
- That high-scoring countries produce better realised user outcomes than lower-scoring countries.
- That the public meaning of a label such as User-Driven is externally valid.

The problem is not weak validation of the capability evidence. The problem is that rigorous capability validation is still not external validation of realised human outcomes. Internal reliability, correlation, factor structure and sensitivity analysis can strengthen confidence that

the DGI measures digital-government maturity consistently; they cannot by themselves establish that a higher User-Driven score predicts a better realised end-to-end citizen outcome.

KEY DISTINCTION

EVIDENCE VALIDATION IS NOT OUTCOME VALIDATION. STATISTICAL ROBUSTNESS IS NOT EXTERNAL VALIDITY.

4. Findings: where capability stops and human reality begins

The core finding survived positive controls, counter-evidence and cross-country contradiction.

1. The DGI is a credible institutional capability benchmark VERIFIED

OECD's current description is consistent and defensible: the DGI assesses enabling foundations rather than specific service outcomes. The benchmark therefore has genuine analytical value for governments seeking to compare institutional maturity.

WHY IT MATTERS

Criticism should not attack the benchmark for failing to be something OECD now explicitly says it is not.

2. User-Driven contains real outcome-relevant capability VERIFIED

The dimension captures meaningful mechanisms intended to place users at the centre: standards, user engagement, testing, feedback, implementation and monitoring. These are not merely paper commitments.

WHY IT MATTERS

The problem is not emptiness. The problem is inference: outcome-relevant inputs are not the same as observed outcomes.

3. User-Driven is not an observed user-performance ranking VERIFIED

The dimension does not directly observe whether citizens complete representative end-to-end journeys, how much effort they expend, how failures are recovered, or whether the final outcome is usable and durable.

WHY IT MATTERS

A country rank on User-Driven cannot be treated as a rank on realised user experience without an additional validation layer.

4. OECD itself confirms the capability-to-delivery gap VERIFIED

The 2026 Digital Government Outlook says the challenge has shifted from building foundations to delivering tangible impact. It documents inconsistent application of service standards,

fragmented handovers, incomplete joined-up delivery and limited measurement of actual user burden.

WHY IT MATTERS

The review's central concern is not external speculation; it is independently corroborated by OECD's own later evidence.

5. User burden remains systematically under-observed VERIFIED

Only 10 of 36 OECD countries report standardised mechanisms or guidance for measuring transaction costs imposed on users. OECD defines these costs broadly: time, steps, documents, handovers and repeat contact when something goes wrong.

WHY IT MATTERS

A benchmark can observe mature capability while missing substantial citizen cost.

6. Digital availability can conceal end-to-end failure STRONGLY SUPPORTED

A portal, digital identity, online form or automated component can work correctly while the total journey fails because another organisation, physical channel, evidence requirement or recovery step breaks the chain.

WHY IT MATTERS

Component success should never be reported as equivalent to journey success.

7. The citizen can become the system integrator STRONGLY SUPPORTED

Where agencies do not carry information, responsibility or recovery across institutional boundaries, citizens perform the integration themselves: finding appointments, obtaining proofs, moving between channels, repeating information and coordinating organisations.

WHY IT MATTERS

The missing cost is often not transaction time but integration work transferred to the user.

8. Recovery maturity can be materially lower than initial-service maturity STRONGLY SUPPORTED

Benchmarks and service metrics tend to observe the designed or happy path more easily than failure and recovery. Yet rights, deadlines and access often depend on what happens after something goes wrong.

WHY IT MATTERS

Human-centred performance requires measurement of failure demand, repeat contact, rework and time-to-recovery.

9. Formal entitlement and point-of-service execution can diverge STRONGLY SUPPORTED

Published rules can permit multiple forms of evidence or recovery while local execution applies narrower documentary expectations. The substantive right may exist while the practical ability to prove it becomes the journey.

WHY IT MATTERS

Proofability and reality are different constructs.

10. Country-level capability scores can hide within-country service variance VERIFIED IN PRINCIPLE

A national institutional score compresses heterogeneous journeys, agencies, geographies, channels and user groups. Strong central capability does not imply uniform delivery at every point of service.

WHY IT MATTERS

National rankings need denominator, boundary and distribution awareness.

5. The decisive cross-country test

If User-Driven is allowed to function as a proxy for realised service performance, its country ordering should have a defensible external relationship with independent user evidence.

The OECD Trust Survey provides independent evidence about experienced administrative interactions. It is not a substitute measure for DGI User-Driven and it does not measure the same construct. Its role in this review is narrower and stronger: an external contradiction test. Country inversions do not prove that User-Driven is an incorrect capability measure; they show that institutional user-centred capability does not reliably reproduce independently observed administrative experience, so an outcome-performance interpretation cannot be assumed.

Country	User-Driven	Administrative-service satisfaction	Interpretation
Portugal	0.94	52%	High capability; materially lower user-reported outcome
Korea	0.91	79%	Convergence control
Estonia	0.72	81% / above 80%	Ordering inversion
Chile	0.90	Below 50%	Severe divergence

5.1 Portugal: positive control and contradiction

Portugal prevents the analysis from becoming one-sided. Satisfaction with administrative services rose from 43% in 2023 to 52% in 2025. That is meaningful improvement. Portugal therefore demonstrates that strong capability can coincide with improving outcomes. But a 0.94 User-Driven score still does not mean that Portugal's realised administrative-service outcome is '94%' or that it outranks Estonia's user experience.

5.2 Estonia: the inversion

Estonia's User-Driven score is materially below Portugal's, while its administrative-service satisfaction is above 80%. The independent user evidence therefore does not reproduce the User-Driven ordering.

5.3 Korea: convergence matters

Korea combines a very high DGI composite and User-Driven score with high administrative-service satisfaction. This is important counter-evidence: the benchmark can align with realised

user evidence. The integrity conclusion is therefore not that DGI and outcomes are inversely related. It is that alignment cannot be assumed.

5.4 Chile: high capability, weak realised signal

Chile provides the strongest divergence case in the reviewed OECD material: a high User-Driven score coexists with markedly weaker administrative-service satisfaction. The case reinforces the need for a contradiction mechanism whenever capability and user evidence move in different directions.

CROSS-COUNTRY CONCLUSION

HIGHER USER-DRIVEN DOES NOT IMPLY HIGHER REALISED ADMINISTRATIVE-SERVICE SATISFACTION. CONVERGENCE EXISTS, BUT IT MUST BE DEMONSTRATED — NOT INFERRED.

5.5 European Commission eGovernment Benchmark: service-delivery contradiction test

The European Commission's 2025 eGovernment Benchmark provides a second independent test at a different measurement layer. The redesigned benchmark assesses approximately 100 key public services across nine life events and uses mystery shoppers to observe government websites and service journeys. Its 'digital public services for citizens' KPI measures the share of administrative steps that can be completed fully online for major life events. This is substantially more observational of service delivery than the DGI capability architecture, although it still does not observe the full human chain of access, attempts, handoffs, failure, recovery, elapsed time, burden and final usable outcome.

For the four-country comparison available in official European Commission reporting, the 2025 service-delivery values are Estonia 96/100, Netherlands 89/100, Spain 89/100 and Portugal 84/100. The 2025 OECD User-Driven values order the same countries as Portugal 0.94, Spain 0.73, Estonia 0.72 and Netherlands 0.56. The ordering is therefore materially different: the country ranking based on institutional User-Driven capability does not reproduce country ordering on a substantially more service-observational European measure.

This does not make the European benchmark the 'truth' against which DGI should be scored. It establishes a measurement boundary. Validation at one layer does not establish validity at the next layer. A defensible benchmark ecosystem should therefore distinguish CAPABILITY → OBSERVED SERVICE DELIVERY → EXPERIENCED INTERACTION → END-TO-END JOURNEY → REALISED OUTCOME → HUMAN / ECONOMIC COST.

6. What a national score can make invisible

What is easy to document is not necessarily what matters most to the person completing the journey.

6.1 Observability and formalisation

Institutional capabilities are comparatively observable: strategies exist, standards can be published, governance bodies can be identified, monitoring mechanisms can be documented. Human outcomes are harder: they occur across agencies, channels, locations, edge cases, failure states and time. This creates a structural risk that formalised capability is measured more consistently than the messy reality it is meant to improve.

6.2 Denominator and exclusion

A service can perform well for the users who reach the measured channel while excluding those who cannot obtain an appointment, cannot authenticate, abandon the process, use an intermediary, or recover through an unmeasured offline route. Outcome measurement therefore needs a denominator that includes access failure, abandonment and recovery — not only completed digital transactions.

6.3 Intermediary masking

Where citizens rely on family members, professionals, paid intermediaries or informal assistance to navigate a public process, the eventual completion can conceal the burden transferred outside government. The service may record a successful outcome while the user has incurred significant additional time, money and dependency.

6.4 Geographic and service-topology effects

A country-level score can also hide whether the service architecture itself transfers burden to users through decentralised appointments, physical attendance, fragmented responsibilities or location-dependent execution. These are not arguments for centralisation by default. They are reasons to measure the human cost of the topology chosen.

MEASUREMENT PRINCIPLE

THE RELEVANT QUESTION IS NOT ONLY 'IS THE SERVICE AVAILABLE?' BUT 'WHO PERFORMS THE INTEGRATION REQUIRED TO ACHIEVE THE OUTCOME?'

7. Human reality: an end-to-end mechanism case

One journey cannot rank a country. It can reveal what the benchmark architecture does not see.

The review used an observed EU-citizen registration and recovery journey in Spain as a mechanism case. The case is not presented as representative national evidence. It is used to test whether burdens can exist outside the capability layer: repeated documentary proof, appointment scarcity, cross-organisational dependencies, printing, physical visits, third-party coordination, recovery and elapsed time.

7.1 The observed mechanism

- Substantive financial resources existed, but the required documentary form became a separate problem.
- Health-cover evidence that existed in one context did not automatically satisfy the administrative proof expected in another.
- Bank and insurance interactions became dependencies in a public-administration journey.
- Repeated appointments, travel, printing and coordination were required before the desired administrative outcome was achieved.
- Recovery after loss of documentation required a new administrative path rather than a frictionless reuse of already-known information.
- Third-party effort — including appointment and insurer coordination — formed part of the real completion cost.

7.2 Why the case matters analytically

The case demonstrates the distinction between formal capability and realised completion. Each participating organisation can execute its own step correctly while the citizen still carries the end-to-end integration burden. This is precisely the kind of cost OECD's 2026 transaction-cost chapter says is insufficiently standardised across countries.

7.3 Journey cost

Cost layer	What must be counted
Direct incremental cost	Travel, parking, printing/copies, document fees, attributable overlapping cover or intermediary costs.
Citizen time	Travel, waiting, appointments, calls, evidence gathering, rework, recovery.
Third-party time	Family, professional or intermediary coordination required to complete the outcome.

Failure demand	Repeat contact and work created because the first attempt did not produce the outcome.
Elapsed-time impact	Delay between initial need and usable outcome.
Cognitive burden	Uncertainty, complexity, repeated decision-making and recovery effort; report separately unless valuation is defensible.

BETTER REALITY METRIC

TIME-TO-OUTCOME + FIRST-TIME-RIGHT + RECOVERY BURDEN + TOTAL CITIZEN EFFORT are more informative about lived delivery than portal availability or isolated transaction time.

8. The value of a high score

A ranking is not only information. It can become reputation, legitimacy, strategy and incentive.

8.1 Reputational value

A high OECD ranking is a credible international signal. Portugal's official digital-government portal highlights its third-place DGI position and 86% score as part of the country's 2026 performance baseline. Such communication is understandable: international comparison creates prestige and can support a narrative of digital leadership.

8.2 Political and institutional value

A favourable ranking can validate reform programmes, strengthen the position of responsible institutions, support budget and policy arguments, and provide political actors with a compact proof point. A complex capability architecture is converted into a number and rank that can be repeated in speeches, strategy documents and public communication.

8.3 Strategic value

Where international ranking positions are incorporated into national objectives, the benchmark becomes reflexive: governments do not merely appear in the index; they can organise improvement efforts partly around the index. This is not improper. Benchmarks are intended to stimulate improvement. But once the measure influences the measured system, the benchmark holder must actively distinguish improvement in observable capability from improvement in the intended human outcome.

8.4 Economic value: what can and cannot be claimed

Digital-government leadership can plausibly contribute to country attractiveness, investor confidence, innovation narratives, administrative efficiency and the perceived quality of the business environment. A high OECD ranking can therefore have economic signalling value. This review does not claim that a particular DGI rank causes a quantified amount of foreign investment, GDP growth or financing advantage; the evidence reviewed does not support such a causal number.

MATERIAL CONCLUSION

THE ABSENCE OF A QUANTIFIED EURO EFFECT DOES NOT MAKE THE INCENTIVE IMMATERIAL. POLITICAL, REPUTATIONAL, INSTITUTIONAL AND STRATEGIC VALUE ARE ALREADY SUFFICIENT TO CREATE RATIONAL PRESSURE TO IMPROVE THE SCORE.

8.5 Who benefits?

- National governments: international standing and evidence of reform progress.
- Digital-government authorities: institutional legitimacy and strategic leverage.
- Political leadership: a simple, authoritative success signal.
- Public administrations: justification for capability investment and reform programmes.
- Countries and ecosystems: potential signalling of administrative modernity and digital readiness.
- OECD: benchmark relevance increases as governments cite and target the index.

8.6 The integrity implication

When the measured entities derive value from the result, self-reported evidence is not automatically invalid. It does, however, increase the importance of independent contradiction and outcome controls. The correct response to incentive is governance, not accusation.

9. When a capability score acquires a larger meaning

Authority can preserve accuracy at the source while meaning expands downstream.

MEANING CHAIN

EVIDENCE → CONSTRUCT → SCORE → LABEL → RANK → OECD NARRATIVE → NATIONAL COMMUNICATION → POLICY / PUBLIC ASSUMPTION

9.1 The authority effect

OECD is an authoritative international institution. That authority is part of the value of the benchmark. It also means that a compact label and rank can travel further than the technical caveats that define the construct. A national communication can accurately report a third-place DGI result while a reader reasonably hears 'third-best digital government' or 'third-best user-centred service performance'.

9.2 Why User-Driven is especially exposed

The phrase User-Driven is close to an outcome claim in ordinary language. It does not naturally signal 'maturity of governmental capabilities intended to enable user-driven transformation'. That semantic compression creates a foreseeable interpretation gap even without any actor making a deliberately false statement.

9.3 Claim distance

The more steps between raw evidence and public claim, the greater the need to preserve the construct boundary. The report therefore treats evidence–claim distance as an integrity property of the benchmark ecosystem, not merely a communications issue.

INTEGRITY RULE

THE GREATER THE AUTHORITY OF THE SCORE, THE SMALLER THE TOLERABLE DISTANCE BETWEEN WHAT IS MEASURED AND WHAT THE SCORE IS ALLOWED TO PROVE.

10. Governance and independence

Peer-government benchmarking is legitimate — but it is not the same as independent observation of people.

OECD develops the indices with Member countries through the Working Party of Senior Digital Government Officials and the Public Governance Committee. Participating administrations supply evidence and review their data; OECD validates evidence and controls the methodology. This is a legitimate peer-government model and should not be misdescribed as countries simply scoring themselves.

10.1 Role concentration

The same institutional ecosystem participates in defining the benchmark, supplying much of the evidence, receiving the score and benefiting from the comparison. OECD provides an independent secretariat and validation role, but the architecture remains participant-intensive.

10.2 Verification asymmetry

Institutional claims are comparatively easy to verify through documents, policies and systems. Negative human outcomes are dispersed across millions of journeys and often require separate observation. Without a mandatory outcome contradiction layer, the architecture is structurally better at validating the presence of capability than at detecting where capability has failed to become delivery.

10.3 The missing gate

The review did not identify an equivalent mandatory scoring gate in which independently observed end-to-end human outcomes can contradict, qualify or suspend a User-Driven country ranking before publication. OECD uses outcome evidence elsewhere — including its Trust Survey and Digital Government Outlook — but those datasets are not integrated as an external-validity gate for the User-Driven ranking itself.

GOVERNANCE CONCLUSION

THE CURRENT ARCHITECTURE IS SUFFICIENTLY INDEPENDENT FOR A CAPABILITY BENCHMARK, BUT NOT SUFFICIENTLY OUTCOME-INDEPENDENT TO SUPPORT A PERFORMANCE INTERPRETATION OF USER-DRIVEN.

11. Mission, purpose and the human beneficiary

A people-centred ambition must ultimately be tested where people experience government.

OECD's digital-government work explicitly aims at coherent, human-centred transformation and tangible impact for people and businesses. That creates a legitimate normative baseline for evaluating whether measurement remains aligned with purpose.

11.1 Mission-to-measurement

The DGI measures many of the capabilities plausibly required to achieve human-centred government. Alignment is therefore strong at the capability layer.

11.2 Mission-to-delivery

OECD's 2026 Outlook shows that implementation remains incomplete: standards are not consistently applied, channels are not always joined up and user burden is not systematically measured. Mission-to-delivery alignment is therefore not guaranteed by a high capability score.

11.3 Mission-to-outcome

The beneficiary is ultimately the person or business that needs an outcome, not the policy document, portal or benchmark participant. If the benchmark is publicly interpreted as evidence of human-centred performance, the human outcome must become part of the validation chain.

PURPOSE TEST

IF PEOPLE ARE THE STATED BENEFICIARY, THEIR REALISED OUTCOME CANNOT REMAIN AN OPTIONAL EXTERNAL DATASET WHEN THE RANKING IS GIVEN A USER-PERFORMANCE MEANING.

12. What the DGI is good for — and what it should not be used for

The strongest remedy preserves what works.

12.1 Appropriate uses

- Comparing institutional digital-government maturity and enabling foundations.
- Identifying policy, governance, platform, data and user-centred capability gaps.
- Supporting peer learning and reform priorities.
- Tracking the development of government-wide digital capability over time.
- Stimulating investment in foundations that are plausibly necessary for better outcomes.

12.2 Uses that require independent outcome evidence

- Claiming that one country's citizens experience better digital public services than another's.
- Treating User-Driven rank as a league table of realised user-centred performance.
- Using a high score as proof that end-to-end journeys are simple, accessible or low-burden.
- Using the ranking as proof that failure and recovery work well.
- Inferring that groups excluded from measured channels experience the same quality as measured users.

12.3 The correct two-layer model

PROPOSED ARCHITECTURE

LAYER A — CAPABILITY: DGI foundations, governance, practices and maturity.

LAYER B — REALISED OUTCOME: representative journeys, access, completion, satisfaction, burden, failure, recovery, equity and time-to-outcome.

DO NOT COLLAPSE THE TWO LAYERS INTO ONE SCORE. MAKE DIVERGENCE VISIBLE.

13. Publication decision: what should happen to the list now?

The conclusion must change the treatment of the ranking, not merely add another caveat.

13.1 Options considered

Option	Assessment	Decision
Keep as is	Does not address foreseeable outcome interpretation or absence of external outcome validation.	REJECT
Add a footnote/disclaimer	Improves technical accuracy but leaves the User-Driven label and comparative rank carrying broader meaning.	INSUFFICIENT ALONE
Relabel User-Driven	Materially reduces semantic overreach when the label explicitly signals capability and comparative use is paired with independent delivery/outcome evidence.	RECOMMENDED
Publish capability + outcome side by side	Preserves DGI value and exposes divergence.	RECOMMENDED
Discontinue current standalone/performance-like presentation	Addresses the unsupported performance-like interpretation while preserving the underlying capability measure; suspension is one possible transition mechanism.	RECOMMENDED NOW
Withdraw the entire DGI	Disproportionate: the capability benchmark itself remains useful and defensible.	REJECT

13.2 Final publication judgement

DECISION
THE CURRENT STANDALONE/PERFORMANCE-LIKE PRESENTATION OF THE USER-DRIVEN COUNTRY RANKING SHOULD NOT CONTINUE IN ITS PRESENT FORM. RETAIN THE CAPABILITY MEASURE, LABEL IT EXPLICITLY AS CAPABILITY, AND PAIR COMPARATIVE USE WITH INDEPENDENT SERVICE-DELIVERY AND HUMAN-OUTCOME EVIDENCE.

13.3 Why a disclaimer alone is not enough

The integrity problem is not confined to a missing sentence in the methodology. It arises from the combination of an intuitive label, a country ranking, OECD authority, national prestige and downstream use. A technical caveat can coexist with a public interpretation that still exceeds the evidence. The remedy must therefore change the presentation architecture, not only the fine print.

13.4 Conditions for republication

- Rename or define the dimension so its capability nature is unmistakable.
- Add independent outcome measures covering representative end-to-end journeys.
- Include access failure, abandonment, handoffs, repeat contact, recovery and total citizen burden.
- Use independent user evidence as a contradiction trigger when capability and outcome diverge.
- Publish the capability score and outcome evidence side by side rather than implying equivalence.
- Document denominator and exclusion boundaries.
- Disclose where country-reported evidence is used and how it is independently checked.
- Provide downstream communication guidance that prevents the rank being described as realised user-performance leadership.
- Reassess countries where large capability/outcome divergence is observed before promoting comparative User-Driven claims.

13.5 Transition

OECD does not need to abandon years of work. It can preserve the DGI as the capability layer, discontinue the current standalone/performance-like User-Driven presentation in its present form, make the capability boundary explicit, and pair comparative use with independent service-delivery and human-outcome evidence. That would increase rather than reduce the value of the index.

14. Recommended redesign

The solution is not less measurement. It is better separation of capability, delivery and outcome.

01. Preserve the DGI capability index

Keep the six-dimension capability architecture where it remains valid, but state the construct boundary on every ranking surface.

02. Replace the current User-Driven ranking presentation

Retain the underlying capability measure, but discontinue its current standalone/performance-like country-ranking presentation in its present form. Use an explicit capability label such as User-Centred Capability / User-Driven Foundations and pair comparative use with independent service-delivery and outcome evidence.

03. Create an independent Human Delivery & Outcome companion

Measure representative end-to-end journeys across life events, channels, geographies and user groups.

04. Measure total citizen burden

Standardise time, steps, documents, handoffs, travel, repeat contacts, intermediary effort and failure demand.

05. Make recovery visible

Track first-time-right, abandonment, repeat contact, time-to-recovery and final usable outcome.

06. Build contradiction into governance

A high capability score plus weak independent outcome evidence should automatically trigger review and explanatory publication.

07. Separate evidence classes

Show what is government-reported, independently observed, administrative-data based and user-reported.

08. Publish uncertainty and boundaries

State what population, channels, services and time periods each claim covers.

09. Control authority echo

Provide approved language for OECD and national communication that preserves the construct boundary.

10. Monitor ranking incentives

Disclose when DGI position is adopted as a national target and examine whether improvement in score is accompanied by improvement in human outcomes.

11. Protect positive learning

Use countries such as Korea and improving Portugal as learning cases where capability and outcomes align or move together.

12. Investigate divergence, do not average it away

Treat Estonia, Chile and other inversions as information about the measurement system rather than inconvenient noise.

15. Counter-evidence, boundaries and what this report does not claim

A hard conclusion is only credible when its limits are equally explicit.

Portugal is not evidence of failure

Portugal's satisfaction improved materially. The report uses Portugal both as positive control and as evidence that capability level and realised outcome level are not interchangeable.

Korea is a convergence control

High DGI capability can coincide with strong user-reported outcomes. This prevents the conclusion from becoming 'high DGI means bad services'.

Satisfaction is not the same construct as DGI

The comparison is used to test proxy interpretation, not to replace the DGI with a satisfaction league table.

The Spain journey is not national proof

It demonstrates mechanisms and missing cost layers. It is not a representative sample of Spanish public services.

Economic value is bounded

The report identifies political, reputational, institutional and strategic value and plausible economic signalling. It does not assign unsupported causal monetary gains to rank positions.

No corruption finding

Incentives, role concentration and reflexivity are governance risks. The review does not find that OECD or participating governments manipulated the index.

No claim that OECD ignores outcomes everywhere

OECD has extensive outcome-oriented work. The finding is narrower: those outcome datasets are not an equivalent mandatory external-validity gate for the User-Driven ranking.

16. Final conclusion and advice to OECD

A useful capability benchmark should not be allowed to become a stronger outcome claim than its evidence can carry.

The OECD Digital Government Index is not a failed benchmark. It is a valuable benchmark being asked — through its most intuitive label and the authority of ranking — to carry more meaning than its measurement architecture has established.

The research objective was to determine whether the DGI measures genuinely user-centred public-service performance or primarily institutional and policy capability. The evidence now supports a clear answer: the DGI primarily measures enabling institutional capability. User-Driven contains substantive practices that can contribute to human-centred delivery, but it is not an independently validated observation of realised end-to-end user performance.

That distinction is not theoretical. OECD Trust Survey evidence does not reproduce the User-Driven country ordering, and the European Commission's substantially more service-observational 2025 eGovernment evidence also produces a materially different ordering across Estonia, the Netherlands, Spain and Portugal. Neither external source is a substitute measure for User-Driven, and neither is a complete end-to-end human-outcome benchmark. Together they demonstrate why validity cannot be carried automatically from institutional capability to observed delivery, experienced interaction or realised outcome. OECD's own 2026 Outlook reinforces the same boundary: governments still struggle to convert foundations into joined-up, reliable delivery and standardised measurement of the burden imposed on users remains rare.

At the same time, the ranking has value. Countries can use it as international proof of progress, leadership and reform maturity. Institutions gain legitimacy. Political actors gain a compact success signal. Rankings can become strategic objectives. Economic signalling is plausible even where direct causal financial effects are not quantified. These benefits create rational incentives to improve the score and amplify its meaning. That makes the absence of an independent human-outcome gate a material integrity issue.

The appropriate remedy is therefore neither complacency nor destruction. OECD should preserve the DGI as a capability benchmark. The current standalone/performance-like presentation of the User-Driven country ranking should not continue in its present form. The capability measure should be retained, its construct explicitly labelled as capability, and comparative use paired with independent service-delivery and human-outcome evidence. Suspension may be used as a transition mechanism, but the evidence supports the required change in presentation and validation architecture rather than one exclusive administrative solution.

FINAL ADVICE
1. KEEP the DGI as a capability benchmark.
2. DISCONTINUE the current standalone/performance-like User-Driven ranking presentation in its present form.
3. RELABEL the construct explicitly as capability.
4. PAIR comparative use with independent service-delivery and human-outcome evidence.
5. BUILD contradiction, denominator, recovery and citizen-burden evidence into the benchmark ecosystem.

This recommendation is demanding because the benchmark matters. A strong international institution should be able to distinguish between the capabilities governments say and show they possess, the systems those capabilities create, and the reality people finally experience. Making that distinction visible would not weaken the OECD Digital Government Index. It would make it more trustworthy, more useful and more aligned with the human-centred transformation it is intended to support.

BETTER BENCHMARKS DO NOT MERELY REWARD WHAT GOVERNMENTS CAN SHOW.
THEY HELP GOVERNMENTS SEE WHAT PEOPLE ACTUALLY EXPERIENCE.

17. Primary sources

The report prioritises OECD and official government sources.

[1] OECD (2026). Digital Government Outlook 2026 — Spain country note. Official source — DGI construct boundary and 2025 scores.

[2] OECD (2026). Digital Government Outlook 2026 — Building human-centred and proactive government services. Official source — Implementation gaps, joined-up delivery and transaction-cost measurement.

[3] OECD (2026). Digital Government Outlook 2026 — Executive summary. Official source — Shift from foundations to tangible impact.

[4] OECD (2026). Digital Government Outlook 2026 — Full report / foreword. Official source — Foundations versus real impact for people and businesses.

[5] OECD (2026). OECD Survey on Drivers of Trust in Public Institutions 2026 — Portugal. Official source — Administrative-service satisfaction 43% to 52%.

[6] OECD (2026). OECD Survey on Drivers of Trust in Public Institutions 2026 — Korea. Official source — Administrative-service satisfaction 79%.

[7] OECD (2026). OECD Survey on Drivers of Trust — day-to-day interactions. Official source — Cross-country user evidence, including Estonia and Chile.

[8] OECD (2026). Digital Government Outlook 2026 — Digital governments at a turning point. Official source — 2025 DGI results and dimensions.

[9] digital.gov.pt (2026). Portugal — Indicators: Where We Stand in 2026. Official source — Official use of third-place DGI rank and 86% score.

[10] OECD (2024). 2023 OECD Digital Government Index: Results and key findings. Official source — Prior DGI architecture and methodological baseline.

[11] OECD (2026). Digital Government Outlook 2026 — Korea country note. Official source — Country score context and development through E-Leaders.

[12] OECD (2026). Digital Government Outlook 2026 — Chile country note. Official source — Chile DGI context and capability framing.

[12] OECD (2025). Methodology for the OECD Digital Government Index, Government at a Glance 2025. Official source — questionnaire architecture, evidence validation, country confirmation and statistical-validation framework.

[13] OECD (2025/2026). DGI 2025 Data Explorer. Official source — systematic supporting-document requirement, Secretariat validation/fact-checking and exclusion of unsupported responses.

[14] European Commission (2025). eGovernment Benchmark 2025. Official source — redesigned service-delivery framework, approximately 100 key public services, nine life events and mystery-shopper observation.

[15] European Commission / State of the Digital Decade country reporting (2025 data). Official source — digital public services for citizens: Estonia 96, Netherlands 89, Spain 89, Portugal 84.

17.1 Evidence-status discipline

Statements in the report are deliberately bounded. VERIFIED is used where primary or strongly triangulated evidence establishes the proposition. STRONGLY SUPPORTED is used where the mechanism is well supported but the available evidence does not justify universalisation. Observed journey evidence is never silently converted into national prevalence. Plausible economic effects are not converted into causal monetary claims without evidence.

Appendix A. Material evidence map

Every material conclusion in the report has a visible destination.

Evidence cluster	Material conclusion	Report destination	Basis
Capability construct	DGI measures enabling foundations, not specific-service digitalisation	§2–4	OECD 2026
Outcome non-equivalence	User-Driven ordering not reproduced by independent user evidence	§5	OECD DGI + Trust Survey
Burden gap	Only 28% standardised transaction-cost measurement	§4, §6–7	OECD DGO 2026
Journey integration	Citizen can carry cross-organisational handoffs and recovery	§6–7	OECD DGO + mechanism case
Observability / denominator	National capability score compresses access, abandonment and local variance	§6	Construct analysis + OECD service evidence
Score value	Rank creates reputational, political, institutional and strategic value	§8	Official national communication + downstream use
Economic signalling	Plausible but not causally quantified	§8	Bounded inference
Authority expansion	Label/rank can acquire broader public meaning	§9	Construct-to-communication trace
Governance independence	Peer-government evidence architecture differs from independent outcome observation	§10	OECD governance architecture
Mission alignment	Human-centred purpose requires human-outcome evidence when claims reach outcome level	§11	OECD stated purpose + DGO
Publication remedy	Current User-Driven presentation requires suspension, relabelling or outcome validation	§13–16	Synthesis

Observed service delivery	DGI User-Driven ordering is not reproduced by a substantially more service-observational European measure	§5	European Commission eGovernment Benchmark / State of Digital Decade 2025

Appendix B. Publication language guardrails

Claims should remain as strong as the evidence — and no stronger.

Use	Avoid
The DGI is a benchmark of institutional digital-government capability and enabling foundations.	The DGI proves which country delivers the best digital public services to citizens.
User-Driven captures meaningful governmental practices intended to put users at the centre.	A higher User-Driven score proves a better realised user experience.
High rankings create reputational, political, institutional and strategic value.	A one-place DGI increase causes a specified amount of GDP or investment without causal evidence.
OECD applies substantial evidence and statistical validation; that architecture does not establish external validity for a realised-outcome interpretation.	OECD performs weak or no validation.
Participant involvement creates a governance need for independent outcome checking.	Countries simply score themselves or the index is corrupt.