

OECD Public Governance

INNOVATING IN
BULGARIA'S PUBLIC
ADMINISTRATION:
**A PRACTICAL GUIDE
TO DESIGNING AND
RUNNING POLICY PILOTS**

Strengthening the capacity to design and
monitor public policies



Funded by
the European Union

This document was produced with the financial assistance of the European Union. The views expressed herein can in no way be taken to reflect the official opinion of the European Union.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

This document was produced by the OECD for internal use by the Bulgarian administration. Mariana Lavin and Josefina Vargas provided methodological and design support throughout the development of this document. The OECD thanks the Council of Ministers' Administration for its leadership throughout the project, as well as the teams across the Bulgarian public organisations that participated in the pilots, and contributed evidence, insights, and feedback that informed the development of this toolkit.

Contact

Joaquín Collao Jul joaquin.collao@oecd.org

Angela Hanson angela.hanson@oecd.org

Sara Kovacheva sara.kovacheva@oecd.org



HOW TO USE THIS GUIDE

This guide is a practical toolkit for public servants and managers who want to design and run innovative policymaking pilots. It is aimed primarily at strategic planning units and policy functions in line ministries and agencies of Bulgaria's central government, and it assumes no prior experience with design thinking, prototyping, or innovation methods. If you are unsure whether this guide addresses your situation, see Box 1 before you begin.

The guide is organised around the Scan-Pilot-Scale framework, a structured approach to testing new policy ideas before committing to full implementation. Each section introduces a phase of the framework, explains its purpose, and offers a set of practical tools. Each tool follows the same format: why it matters, what it helps you do, how to use it step by step, and what to expect in practice.

You do not need to use every tool in every pilot. Think of this guide as a menu: choose the tools that best fit your context, your resources, and the question you are trying to answer. The ten tools included here are drawn from the methods used in three pilots implemented in Bulgaria between 2025 and 2026. These pilots, conducted as part of the project Strengthening the Capacity to Design and Monitor Public Policies, supported by the European Commission's Technical Support Instrument¹, were implemented with technical guidance from the OECD and in collaboration with the Council of Ministers' Administration (CoMA).

The guide builds on the experience of the three pilot initiatives implemented through the project: responsible AI experimentation led by the Ministry of Justice, cross-generational mentoring led by the National Employment Agency, and personalised learning pathways led by the Institute of Public Administration. With the support of CoMA, these pilots brought together teams from seven public organisations through a structured sequence of Scan, Pilot, and Scale activities. Their participation generated the practical evidence, examples, reflections, and lessons that underpin this toolkit, helping translate innovation methods into guidance adapted to the context of Bulgaria's public administration.

Boxes throughout the guide highlight examples from Bulgaria's own pilots and from international practice, to show what these methods look like in real institutional contexts. All tool canvases referenced throughout this guide are available in full size in Annex A at the end of this guide.

¹ The project was funded by the European Union via Technical Support Instrument, and implemented by the OECD, in cooperation with the European Commission.

Box 1. Before you begin

WHAT IS PUBLIC SECTOR INNOVATION?

The OECD defines public sector innovation as something new or novel, implemented and intended to positively impact a specific context, through improved outcomes, greater efficiency, or increased public trust in government (OECD, 2017). It can serve four objectives: reaching strategic goals (mission-orientation); preparing for future challenges (anticipation); improving current systems (enhancement); and responding to evolving citizen needs (adaptation) (OECD, 2022).

In Bulgaria, innovative efforts have emerged across diverse policy areas, including economic development, digital service delivery, and administrative simplification. For more information, you can consult the OECD report on Strengthening the Innovative Capacity of the Public Sector of Bulgaria (OECD, 2024).

IS THIS THE RIGHT GUIDE FOR YOUR CHALLENGE?

IF YOU ARE...	THIS GUIDE WILL HELP YOU...
Working on a policy or service challenge where you know something needs to change, but you are not yet sure what the right solution looks like.	Structure your thinking before you build anything, using problem framing, stakeholder mapping, and blueprint tools to understand the challenge deeply before committing to a direction.
Tasked with piloting a new approach in your public organisation and unsure how to design, test, or document it.	Move from a broad idea to a testable, evidence-grounded pilot design with a clear scope, a realistic timeline, and a plan for what you will learn.
Preparing to propose an innovative initiative to a manager or senior leader and needing to build a credible case.	Develop the evidence, the theory of change, and the communication tools needed to present a compelling, honest case for moving forward.
Running a pilot that has produced early results and now needs to be evaluated, communicated, or taken to a wider audience as a basis for scaling.	Design an evaluation framework that captures what worked and why and develop a scaling plan and pitch that can carry the initiative beyond its original scope.
A manager or senior leader who wants to support your team in working more innovatively but is unsure what that looks like in practice.	Understand what each phase of an innovative policymaking process requires from leadership, in terms of time, resources, endorsement, and the kind of culture that makes experimentation possible.

Source: OECD

This guide closes with a section on mainstreaming the practice of innovation, describing the behaviours and capabilities that help public servants and their leaders turn a single pilot into a more lasting institutional practice. Building these behaviours matters beyond any individual pilot project or specific innovation initiative. OECD research shows that innovative capacity is one of the most significant factors explaining why some public organisations perform better than others over time, delivering more effective policies, stronger public trust, and better outcomes for the people they serve (OECD, 2025).

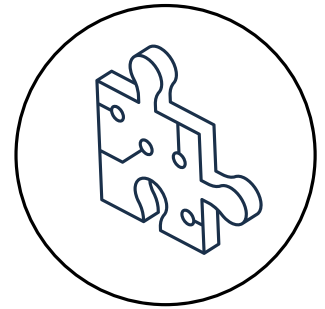
The guide has been developed by the OECD Observatory of Public Sector Innovation (OPSI), part of the OECD Public Governance Directorate. Since 2011, OPSI has worked with governments, academia, industry, and civil society around the world to put innovation at the heart of government: identifying new practices at the edge of public administration, understanding how innovation can become part of the day-to-day work of public servants, and providing trusted guidance to help governments achieve better outcomes for their people. The methods in this guide draw on that global experience and have been adapted for the Bulgarian public administration context.

TABLE OF CONTENTS

1 - Why innovative policymaking pilots matter	07
2 - The Scan-Pilot-Scale approach The three phases at a glance	11
3 - Key tools for innovative policymaking Scan phase Pilot phase Scale phase	13 16 24 46
4 - Mainstreaming the practice of innovation: Skills that make the difference If you are a public servant If you are a manager or leader	57 58 59
References	60
Annex A. Tool Canvases	62

01

WHY INNOVATIVE
POLICYMAKING
PILOTS MATTER



Public administrations today are operating in an increasingly complex and fast-moving environment, where established tools and processes must often be adapted to respond to emerging challenges. Public administrations today are operating in an increasingly complex and fast-moving environment, with tools and processes designed for a different era. Traditional policy development cycles, long planning horizons, extensive approval chains, and limited room for iteration, struggle to keep pace with the speed at which challenges emerge and circumstances change.

Innovative policymaking is a complementary set of tools and methods that can help strengthen, modernise and complement traditional approaches to policy design. It introduces structured ways to test ideas before committing to them at scale. It invests in understanding problems deeply before proposing solutions. And it builds the capacity to learn from what works, and what does not, rather than only from what was planned.

The OECD Declaration on Public Sector Innovation (2019) recognises this challenge directly, calling on governments to pursue government innovation as a core function of the public administration, not a peripheral activity reserved for specialised innovation teams, but an embedded practice across all public organisations. The Declaration identifies the ability to experiment, learn, and adapt as central to governments' capacity to respond to complex and uncertain environments (OECD, 2019). For public servants working in strategic planning functions, this translates into three concrete benefits, set out in Box 1.1 below.

Box 1.1. Three benefits of running a policymaking pilot

- 1. Generate evidence before you commit:** A pilot allows you to test whether a proposed solution actually works in your institutional context, before it is written into a strategy, approved by leadership, or rolled out at scale. The evidence it produces is more credible and more specific than desk research alone.
- 2. Surface what you cannot see from the inside:** Operational constraints, user barriers, and coordination gaps rarely appear in planning documents. A pilot brings them to the surface early, when they are still easier to address, rather than after full implementation has begun.
- 3. Build institutional knowledge and capability:** Running a pilot develops skills in problem framing, user research, prototyping, and evaluation that strengthen your team's capacity to design better policies in the future, not just the one you are currently working on.

Source: OECD

As Box 1.1 illustrates, piloting means having tools to test a new approach on a small scale before recommending it for broader implementation in a strategic document, being able to show evidence that a proposed solution addresses a real need, and being equipped to communicate what you learned, both what worked and what needed to change.

Pilots are an integral part of sound policy design. A well-designed pilot produces knowledge: it surfaces assumptions that need to be tested, reveals operational constraints that were not visible in the planning stage, and generates user insights that improve the final design. In Bulgaria, three projects implemented under the project Strengthening the Capacity to Design and Monitor Public Policies in Bulgaria (2024-2026) demonstrated exactly how the process of designing and running pilots produced concrete evidence, institutional learning, and more realistic implementation plans..

Knowing why pilots matter is only the first step. Choosing the right challenge to pilot is equally important, and equally worth thinking through carefully before you begin. A well-chosen pilot is not only technically feasible and strategically relevant; it is also one that can mobilise the attention, commitment, and collaboration of the public administration around a shared opportunity. Box 1.2 offers simple guidance for assessing whether a challenge is a good candidate for piloting.

This guide distils that experience into a set of tools that any team in Bulgaria's public administration can use, whether running a first pilot or building on earlier experience. If your work is more directly focused on the design and delivery of public services, the OECD Public Service Innovation Booster offers a complementary resource specifically developed for that context (See Box 1.3 below).



Box 1.2. How to choose the right challenge to pilot

Choosing the right challenge is one of the practices that sets a pilot up for success. The three criteria below, drawn from the OECD's experience selecting pilots in Bulgaria's public administration, offer a practical way to assess whether a challenge is well-positioned for piloting and to identify what conditions need to be in place before you begin.

1. Strategic alignment and expected impact: Does the challenge address a real and well-defined policy problem? Is it relevant to your institution's priorities and, where applicable, to national or EU strategies? Will a successful pilot produce measurable improvements, in efficiency, service quality, or public outcomes, that are worth the investment?

2. Implementation feasibility: Is there a clear institutional lead with the mandate and capacity to run the pilot within a realistic timeframe? Are the technical, legal, and data requirements manageable with available resources? A good pilot is ambitious enough to be meaningful but scoped tightly enough to be deliverable.

3. Long-term potential: If the pilot works, can it be sustained and scaled? Is there a plausible pathway to integrating the solution into normal institutional practice, or expanding it to other organisations or policy areas? Pilots that are designed as one-off experiments from the start rarely produce lasting change.

The strongest candidates for piloting show some strength across all three criteria, though few challenges will score equally well on every dimension from the start. What matters is having enough in each area to make the pilot feasible: a clear institutional lead, a challenge that is realistically addressable within available resources, and at least a plausible pathway toward broader adoption. Where gaps exist, the question is not necessarily whether to proceed, but what conditions need to be strengthened or what adjustments to the scope would make the pilot more viable before you begin.

Source: OECD

Box 1.3. Complementary resource: OECD Public Service Innovation Booster

This guide is designed to complement the OECD Public Service Innovation Booster, a toolkit that supports public sector teams in developing innovative solutions to public service challenges. The Booster uses a structured sprint methodology combining tools such as problem framing, stakeholder mapping, and prototyping. The tools in this guide draw on and adapt several Booster methods for use in Bulgaria's public administration context.

For more information, visit www.oecd-opsi.org

Source: OECD.

02

THE SCAN-PILOT-
SCALE APPROACH

The Scan-Pilot-Scale framework provides a structured pathway for testing and scaling innovative policy ideas. It breaks down what can feel like an extensive process, how do we actually try something new?, into three manageable phases, each with a clear purpose and a defined set of activities.

The three phases at a glance

SCAN

Identify and understand the problem before proposing solutions. Define who is affected, what their needs are, what root causes drive the problem, what already exists, and what constraints any solution must navigate.

PILOT

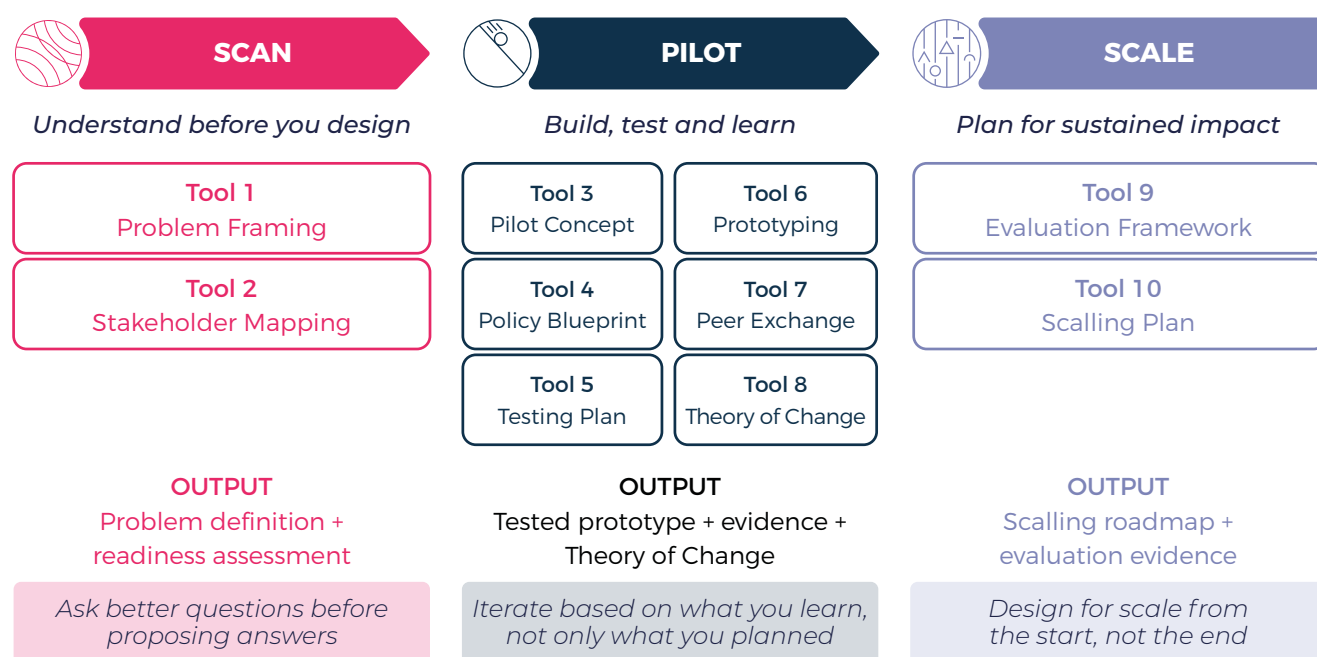
Design and test a small-scale version of your proposed solution. Build prototypes, surface assumptions, gather user feedback, and refine the concept based on what you learn.

SCALE

Plan for broader implementation. Define what evidence you need to justify scaling. Identify what needs to change, in structures, resources, or partnerships, for the solution to work at full scale.

Figure 2.1 illustrates how the three phases connect in practice, showing the tools available at each stage, the outputs produced by each phase, and the iterative logic that runs throughout the process. The tools shown are those covered in this guide; each phase can draw on a much wider range of methods depending on the challenge, the context, and the team's capacity. The framework is not a rigid sequence: teams may move back and forth between phases as they learn, and the insights generated at each stage feed back into the design of the next.

Figure 2.1. Scan-Pilot-Scale approach



Each phase informs and improves the next and the cycle can restart with new challenges

This framework is used in a growing number of countries and organisations. The United Kingdom's AI Opportunities Action Plan (2025) explicitly adopts a Scan-Pilot-Scale approach as the methodology for trialling artificial intelligence across public administration, recognising its value as a disciplined pathway from promising idea to responsible implementation (UK Government, 2025). Chile's GovLab, the government's innovation laboratory, uses an equivalent iterative approach, moving from problem exploration through rapid prototyping to scaled deployment, emphasising user research and continuous learning at each stage (Government of Chile, 2022). The OECD has supported similar approaches across multiple countries, recognising that structured piloting is central to building adaptive and learning-oriented public administrations (Tönurist & Hanson, 2020).

Across all three phases, several principles apply.



Start with the problem, not the solution: It is tempting to jump

straight to a preferred approach, but the strongest pilots emerge from a deep understanding of the challenge and the people it affects.



Work iteratively: Expect to revise your design as you learn more and

build that revision into your plan from the start.



Document as you go:

The knowledge produced by a pilot is only valuable if it is captured and shared. And involve the people who will use or be affected by the policy from the earliest possible stage.

The tools in the following section are organised by phase. Each tool describes where in the Scan-Pilot-Scale journey it is most useful, what it produces, and how to use it in practice.

The tools described in this toolkit represent a selection from a much wider universe of innovation methods available to public sector teams. For those who want to explore beyond this guide, the OECD Toolkit Navigator brings together hundreds of freely available innovation toolkits from across the public, private, academic, and not-for-profit sectors, covering everything from design thinking and behavioural insights to futures and foresight and systems change. See Box 2.1 for more details.

Box 2.1. OECD Toolkit Navigator: Exploring a wider range of innovation methods

The OECD Toolkit Navigator is an available online compendium of innovation toolkits for the public sector, curated by the OECD Observatory of Public Sector Innovation (OPSI) and partners worldwide. It provides access to hundreds of practical resources organised by topic, including public policy, service design, strategic design, behavioural insights, and process facilitation, making it easy to find methods relevant to any phase of your pilot or any challenge you are working on.

Many toolkits in the Navigator are available in adaptable formats that teams can download and modify for their own context. The Navigator also links to government case studies applying these methodologies and connects users with a broader network of public sector innovators.

For more information, visit www.oecd-opsi.org

Source: OECD.

KEY TOOLS FOR INNOVATIVE POLICYMAKING

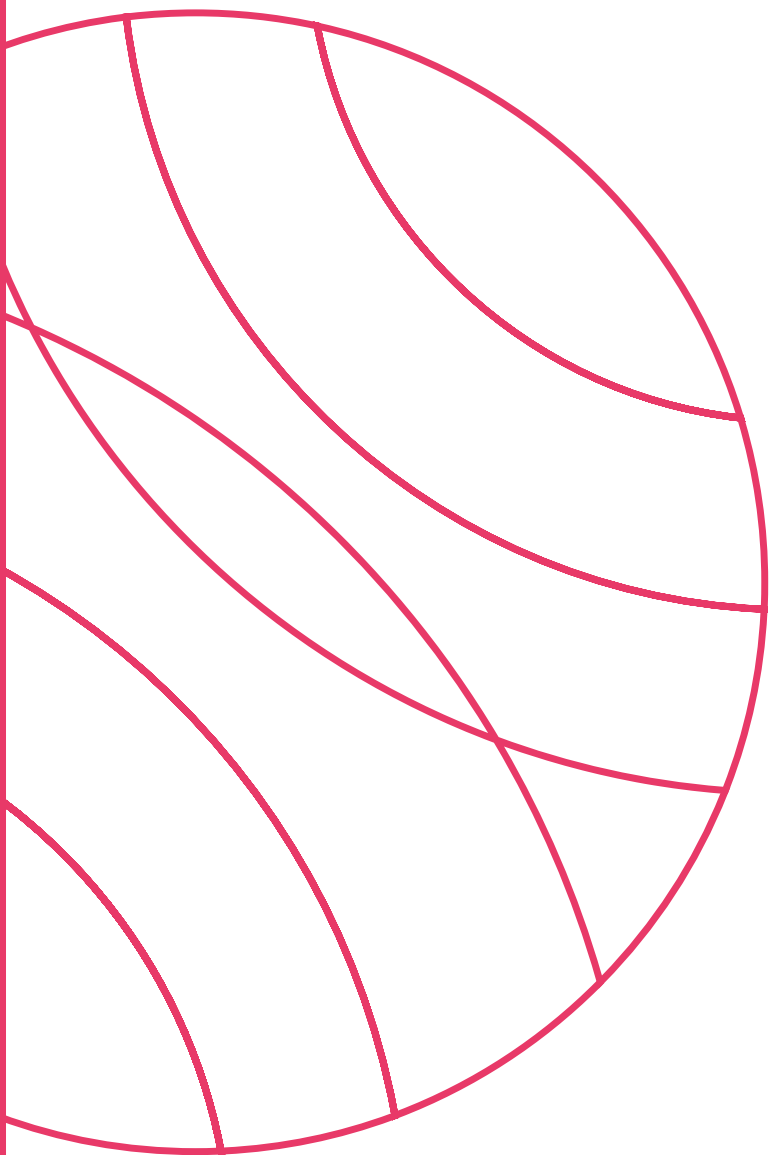
03



KEY TOOLS FOR INNOVATIVE POLICYMAKING

The selected key tools in this section cover the full Scan-Pilot-Scale journey. Based on the experience of carrying out innovative policymaking pilots in Bulgaria, these tools have been identified as priority methods to which all public organisations seeking to run innovative projects should pay particular attention. They are designed to be used flexibly, adapt them to your context, your team's capacity, and the specific challenge you are working on. Each tool can be used in a single working session or developed over multiple workshops or working meetings. These activities can be carried out internally within the pilot team, but they are also encouraged to be participatory wherever possible, involving a broader set of stakeholders, including internal and external users, when feasible.

Throughout the process, you do not need to start from scratch: existing evidence, strategic documents, studies, evaluations, survey data, or previous assessments, are a valuable input at every stage and should be drawn on wherever it is available. The tools in this guide are designed to complement and build on what your public organisations and relevant stakeholders already know.



SCAN PHASE



SCAN PHASE

Before you design anything, understand the problem. The Scan phase is about asking better questions, not proposing answers yet. These two tools help you frame the challenge clearly and assess whether your team has what it needs to run a pilot. Central to this phase, and to the entire Scan-Pilot-Scale journey, is understanding who your policy users are and what they actually need. Box 3.1 introduces the key methods and tools for doing this before you begin. Moreover, Box 3.2 illustrates how the scan phase was applied in practice in Bulgaria and beyond.

MAIN PURPOSE:

Understand the problem deeply before proposing any solution. By the end of this phase, you should have a precise problem definition, a clear understanding of your stakeholder landscape, and an assessment of whether your team is ready to address the pilot's challenge.

WHAT ARE THE KEY BENEFITS OF APPLYING THESE TOOLS?

Tool 1 - Challenge Framing: Ensures your team is solving the right problem, grounded in evidence rather than assumptions, before investing resources in a solution.

Tool 2 - Stakeholder map and feasibility assessment: Prevents surprises mid-implementation by identifying critical actors, potential resistors, and gaps in readiness before you begin.

Tool 1: Challenge framing

Why it matters

Most policy failures begin not with a bad solution but with an incomplete understanding of the problem. Problem framing is the process of moving from a vague sense that something is wrong to a precise, shared definition of what needs to change, for whom, and why. A well-framed challenge gives your team a clear direction and helps you avoid investing resources in solutions that address the wrong issue.

What it helps you do

- Articulate the problem in terms of who is affected and how.
- Distinguish symptoms from root causes.
- Define the boundaries of what your pilot will and will not address.
- Build a shared understanding across team members and stakeholders.

How to use it

1. **Start by drafting the problem.** Write a first version of the problem in one or two sentences using Figure 3.1. The canvas provides a structured sentence to complete (user, need, cause, and consequences) to help you frame your initial problem clearly. Do not aim for precision at this stage; this is a starting point.
2. **Identify who is affected and how.** Ask: who is directly affected by this problem? Who else is affected indirectly? What does the problem cost them in terms of time, money, opportunity, or wellbeing?
3. **Explore root causes.** Ask: why does this problem exist? Identify at least three possible causes. Then, for each cause, ask why it exists. This "5 Whys" approach helps uncover the root of the challenge.
4. **Review what already exists.** Examine relevant strategies, evaluations, studies, or data produced or commissioned by your institution. What has already been tried? Why did previous efforts fall short? Drawing on existing evidence strengthens your problem definition and avoids duplication.
5. **Refine the problem and define the challenge.** Rewrite the problem based on what you have learned. Use the canvas structure to formulate a clear challenge statement. A strong problem definition clearly identifies the affected group, describes the gap or challenge they face, and avoids embedding a solution.
6. **Validate the definition and check existing evidence.** Test the problem and challenge statements with at least one person outside your team. Are they clear? Do they reflect their experience of the issue? At the same time, assess the strength of your evidence: What evidence is available that demonstrates the problem exists?

Common challenges

- Teams often frame problems in terms of the solution they already have in mind ('we need a new platform') rather than the underlying challenge. Keep asking 'why' until you reach a level where the solution is not predetermined.
- Problem definitions can become too broad to be actionable. It helps to specify the scope clearly: which target group, which stage of a process, which institution.
- Getting agreement on the problem definition across team members with different perspectives takes time, but it is worth it since divergent framings could cause conflict later.

Figure 3.1. Problem framing and challenge definition

INITIAL PROBLEM STATEMENT	Complete the sentence: <i>(User/s) need (need) due to (cause), which leads to (consequence).</i>	
WHO IS AFFECTED AND HOW Directly affected: Indirectly affected: Impact (time, cost, opportunity, wellbeing):	ROOT CAUSES (WHY DOES THIS PROBLEM EXIST?) Cause 1: Cause 2: Cause 3: Why? (optional deeper causes)	WHAT ALREADY EXISTS (EVIDENCE AND PAST EFFORTS) Evidence / data / documents: What has been tried: Why it fell short:
REFINED PROBLEM STATEMENT	Complete and improve the sentence: <i>(User/s) need (need) due to (cause), which leads to (consequence).</i>	
CHALLENGE DEFINITION ("HOW MIGHT WE...") Complete the sentence: <i>How might we (action) (object/system of intervention) for (user/s)?</i>	VALIDATION AND EVIDENCE Who tested it: What evidence shows the problem exists and is relevant: Does it reflect their experience: Key feedback:	

Note: The tool canvas in full size is available in Annex A.

Source: OECD

Box 3.1. Understanding your users: a foundation for every phase of the Scan-Pilot-Scale approach

Every policy or service has users: the people who experience the problem, interact with the solution, or are affected by what you design. In a public administration context, users may be civil servants who will use a new tool, citizens who will receive a new service, or managers who will implement a new process. Understanding who they are, what they need, and what barriers they face is not a preliminary step that happens before the real work begins, it is a foundation that runs through every phase of the Scan-Pilot-Scale process.

Investing in user understanding early pays benefits at every subsequent stage. Pilot designs built on assumptions about user needs, and not data and actual experiences, frequently fall short of their intended impact, not because the idea is wrong but because the evidence base was insufficient. The earlier you engage with real users, the cheaper and easier it is to correct your design. During the Scan phase, user understanding helps you frame the problem accurately. During the Pilot phase, it informs your blueprint, shapes your prototype, and guides your testing plan. During the Scale phase, it helps you anticipate how different user groups will experience a broader rollout.

METHODS YOU CAN USE THROUGHOUT THE PROCESS:

- Interviews give you depth. They are good for understanding motivations, barriers, and lived experiences and user journeys that are hard to quantify. Ten to fifteen interviews typically surface the main themes.
- Surveys give you breadth. They allow you to collect quantitative information and test patterns across a larger group and produce data that can be compared before and after implementation.
- Observation, watching users interact with a process or system in real time, gives you unfiltered behavioural data that interviews and surveys can miss.
- Analysis of existing administrative data, evaluations, strategic documents, or previous studies should always be consulted before designing new data collection, and can substantially reduce the research burden at the Scan phase.
- Where possible, combine at least two methods.

TOOLS FROM THE OECD PUBLIC SERVICE INNOVATION BOOSTER THAT CAN SUPPORT YOU:

- The User Map helps you identify the full range of user groups connected to your challenge, distinguishing between external users (citizens, businesses, communities) and internal users (public servants, managers), and between frequent users and extreme users whose uncommon needs can reveal hidden problems and opportunities.
- The User Profile tool supports the development of detailed personas for priority user groups, capturing their demographic characteristics, digital skills, behaviours, expectations, and barriers, not to describe a single individual, but to represent a broader group who share similar characteristics and needs.
- The User Interview guide provides a structured approach to planning and conducting interviews that generate actionable insights rather than general impressions.

These tools are available at www.oecd-opsi.org and can be used at any stage of the Scan-Pilot-Scale journey where you need to deepen your understanding of the people your pilot is designed to serve.

Source: OECD

Tool 2: Stakeholder mapping and feasibility assessment

Why it matters

No pilot succeeds in isolation. Every policy initiative involves people and public organisations, those who will be affected by the change, those who need to approve it, those who can champion it, those who will implement it, and those who might resist it. Stakeholder mapping helps you understand this landscape before you begin, so that your pilot design accounts for the full range of actors involved.

Feasibility assessment adds a second layer: given your stakeholder landscape, your available resources, and your institutional context, is this the right pilot to run right now? Answering this question, even when the answer is 'not yet', is a sign of good feasibility judgement.

What it helps you do

- Identify all groups with a stake in your pilot and distinguish their interests.
- Assess levels of support, neutrality, or resistance across your stakeholder landscape.
- Identify which organisations or individuals need to be involved for the pilot to succeed.
- Evaluate your team's readiness across key dimensions: leadership commitment, staff capacity, resources, and regulatory context.

How to use it

1. **List all relevant stakeholders using Figure 3.2.** Start by identifying the key actors connected to the pilot across four categories:
 - a. **Who is affected:** users and other actors who are directly or indirectly affected by the pilot, including final and intermediate users.
 - b. **Who has authority:** actors with decision-making power over the challenge, the pilot, or the conditions needed for implementation.
 - c. **Who has relevant expertise:** people or organisations with knowledge that can help the team better understand the problem or design the response, including experts from academia, other public organisations, private and third-sector organisations, or international organisations.
 - d. **Who can help implement:** actors whose collaboration is needed to deliver, test, or sustain the pilot.
2. **Assess each stakeholder group.** For each stakeholder, assess their level of interest, likely position, influence, and whether they are critical to the pilot's success. Use simple categories where possible: high or low interest; supportive, neutral, or resistant position; high or low influence; and critical or non-critical stakeholder.
3. **Clarify the role of critical stakeholders.** For each stakeholder whose involvement is essential, ask: what do they need to see or hear to support this pilot?

What concerns might they have? What role should they play? How should the team engage them?

4. **Complete the feasibility check.** Assess whether the pilot has the minimum conditions needed to move forward across five dimensions: senior leadership commitment, sufficient staff time and capacity, access to budget, a regulatory environment that allows experimentation, and access to the users or target groups who need to be involved.
5. **Map next steps.** If significant gaps emerge, identify what would need to change before proceeding. This may involve securing additional support, involving new stakeholders, adjusting the implementation approach, or narrowing the pilot's scope to make it more feasible

Common challenges

- Stakeholder maps tend to undercount resisters. A realistic map should include those likely to push back, why they may do so, and how to engage them, whether directly or indirectly.
- Feasibility assessments are most useful when completed collaboratively with other key stakeholders. Avoid assuming that missing resources will materialise, as this can lead to pilots stalling mid-implementation.
- Find your champions. Identify people at different levels of the public administration who can promote the pilot, explain its value, and help maintain momentum.

Figure 3.2. Stakeholder mapping and feasibility assessment

STAKEHOLDERS									
Who is affected:		Who has authority:		Who has relevant expertise:			Who might be negatively affected:		
STAKEHOLDER ASSESSMENT (FOR EACH KEY STAKEHOLDER/GROUP)					CRITICAL STAKEHOLDERS AND RESISTORS				
Category	Interest	Position	Influence	Critical	Critical stakeholders (essential for success):			Key potential resisters:	
ENGAGEMENT APPROACH (FOR CRITICAL STAKEHOLDERS)			FEASIBILITY CHECK			GAPS AND ADJUSTMENTS			
What do they need to support the pilot:			Senior leadership commitment:			Key gaps identified:			
Likely concerns:			Staff time and capacity:			What needs to change:			
How will you engage them:			Budget availability:			Adjustments to the pilot's challenge scope:			
			Regulatory environment:						
			Access to users / target groups:						

Note: The tool canvas in full size is available in Annex A.

Source: OECD

Box 3.2. What does the scan phase look like in practice?

BULGARIA: NARROWING AN AI CHALLENGE TO A SOLVABLE PROBLEM – MINISTRY OF JUSTICE

As part of the pilots run through the Strengthening the capacity to design and monitor public policies project, supported by the European Commission, the OECD, and the Council of Ministers' Administration, the Ministry of Justice in Bulgaria undertook a pilot on AI experimentation.

When the Ministry of Justice began its pilot, the initial framing was broad: how could AI help the ministry work more effectively? Through the Scan phase, including a cross-institutional survey of 111 civil servants across the Bulgarian public administration and a series of validation sessions, the team progressively narrowed the challenge to a much more precise definition.

The survey revealed that civil servants spent significant time answering repetitive internal procedural questions from colleagues, questions about leave procedures, administrative processes, and regulatory requirements. The problem was not AI adoption in general; it was a specific information bottleneck that AI-powered tools could realistically address.

This precise framing, grounded in evidence rather than assumptions, allowed the team to develop a focused technical specification for an internal AI chatbot assistant rather than a general AI strategy. The process of defining the problem clearly was, in itself, one of the most valuable outputs of the Scan phase (OECD, 2026).

INTERNATIONAL EXAMPLE: NORWAY'S NEW IN NORWAY PORTAL

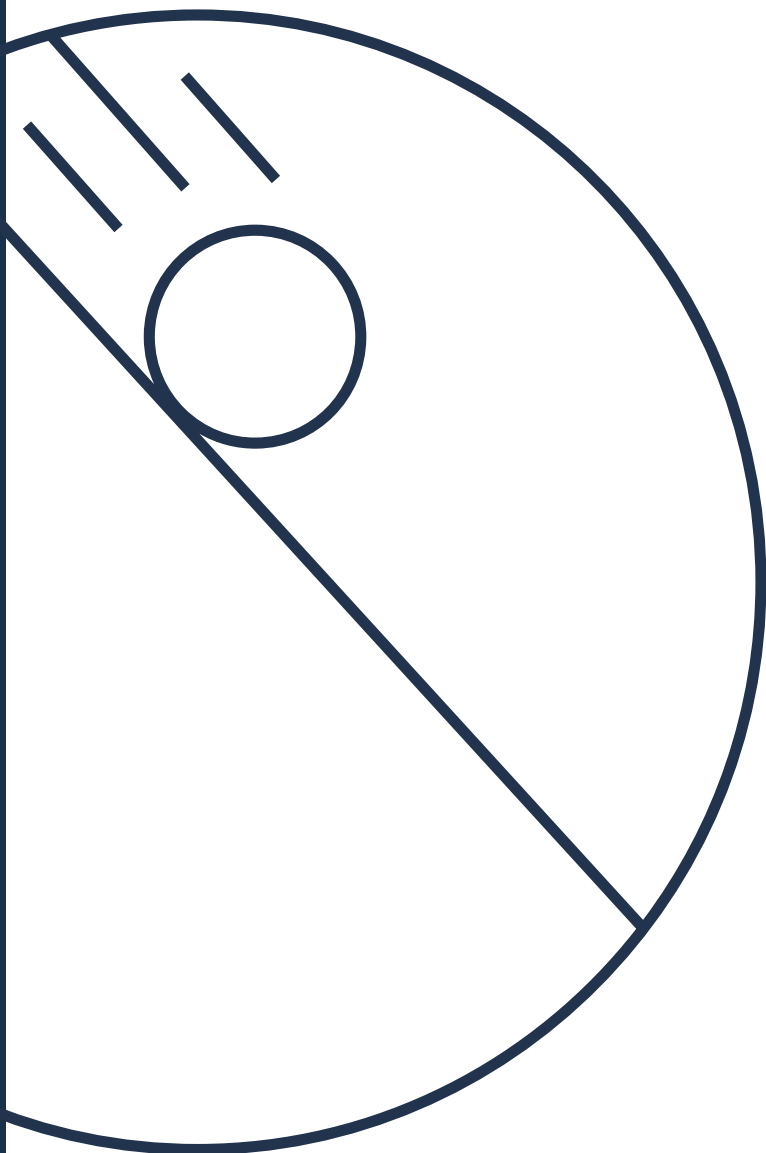
When Norway wanted to improve integration services for newcomers, it did not start by building a new platform. It started by understanding the problem. The Directorate of Integration and Diversity (IMDi), working with the Digitalisation Agency (Digdir), organised a 'Future Gathering' to explore the needs of newcomers and the factors shaping their experiences.

The team analysed 13 driving forces affecting migrants and public services, including digitalisation, changing competence requirements, and polarisation, and developed six future personas representing different user groups: asylum seekers, resettlement refugees, family immigrants, labour immigrants, foreign students, and children in crisis situations.

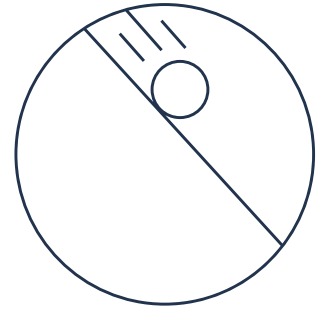
This deep problem framing and stakeholder mapping exercise, completed before any digital solution was designed, produced a precise, evidence-based definition of what newcomers actually needed. It also identified blind spots and areas of low preparedness that shaped the design of the resulting portal.

When an unexpected wave of Ukrainian citizens' arrivals significantly exceeded projections in 2022 and 2023, the portal's design, shaped by this careful Scan phase, allowed it to adapt quickly and serve a much larger and more diverse user base than originally anticipated (OECD, 2024).

Source: OECD based on cited references.



PILOT PHASE



PILOT PHASE

The Pilot phase is where you build, test, and learn. Moving from a well-defined problem and a clear stakeholder landscape, you now develop a structured concept for what to test, map how it would actually work in practice, identify and test the assumptions embedded in your design, and refine the solution through prototyping and peer feedback. By the end of this phase, you will have a tested, evidence-informed pilot design and a clear account of how it is expected to create change. Box 3.5 illustrates how the Pilot phase was applied in practice in Bulgaria and beyond.

MAIN PURPOSE:

Design, test, and refine a small-scale version of your proposed solution. By the end of this phase, you should have a structured pilot concept, a mapped policy journey with surfaced assumptions, evidence from user testing, a strengthened design informed by peer feedback, and a theory of change that articulates how and why the pilot is expected to create impact.

WHAT ARE THE KEY BENEFITS OF APPLYING THESE TOOLS?

- **Tool 3 - Pilot concept:** Translates the insights from the Scan phase into a first structured, testable idea, aligning your team around a shared direction before investing in design and prototyping.
- **Tool 4 - Policy blueprint:** Maps the full policy journey from the user's perspective and the organisation's perspective simultaneously, revealing operational dependencies and surfacing the assumptions that most need to be tested.
- **Tool 5 - Testing plan:** Turns the assumptions surfaced in the Blueprint into a prioritised, structured learning agenda, ensuring that testing is focused on what matters most rather than what is easiest to test.
- **Tool 6 - Low-fidelity prototyping:** Makes your concept tangible and testable before significant resources are committed, generating specific user feedback that improves the design at the lowest possible cost.
- **Tool 7 - Peer exchange and feedback:** Exposes your design to constructive critical challenge from people outside your team, surfacing blind spots and strengthening the solution before implementation begins.
- **Tool 8 - Theory of change:** Consolidates the logic of your pilot into a clear causal account of how your activities are expected to produce outcomes and impact, providing the foundation for evaluation in the Scale phase.



Tool 3: Pilot concept

Why it matters

A pilot concept translates the insights from the Scan phase into a first structured idea of what to test. It is not a final solution, but a working hypothesis about how a specific intervention could address a clearly defined problem and create value.

Without a shared concept, teams often move forward with different interpretations of the problem, the users, or the intended approach. A clear pilot concept creates a common direction and ensures that subsequent design and testing efforts are coherent and focused.

What it helps you do

- Translate a well-defined problem into a clear and testable intervention.
- Articulate who the pilot is for and how it is expected to create value.
- Identify and structure the core elements of the proposed solution.
- Align team members and stakeholders around a shared idea.
- Establish a foundation for prototyping, testing, and iteration.

How to use it

1. **Generate an initial concept.** Build on your problem definition to explore possible solutions. If useful, use structured ideation techniques (e.g. methods from the OECD Public Service Innovation toolkit such as “Crazy 8s” – see Box 3.3) to generate and compare different ideas. Select one concept to develop further.
2. **Describe the solution.** Using Figure 3.3, write a short description of the proposed intervention. Give it a name, define its purpose, and describe in what situations it would be used. A simple visual representation can help clarify the idea.
3. **Define the key features.** Identify the main components of the solution. For each feature, explain why it matters and how it would work in practice at a high level.
4. **Map phases or key activities.** Outline the main stages of how the solution would function. Who is involved at each stage? What tools, systems, or processes are required?
5. **Identify public and stakeholder value.** Ask: who benefits from this solution? What value does it create for different users and stakeholders, and why does that value matter?
6. **Build the case for the pilot.** Draw on existing evidence, research, or user insights. What needs or challenges suggest this solution is worth testing?
7. **Envision expected results.** What changes do you expect to see in the short, medium, and long term? How will you know whether progress is being made?
8. **Communicate the idea clearly.** If you had only 30 seconds to explain the pilot, what would you say? What makes this solution relevant and different from existing approaches?

Common challenges

- Teams may move too quickly into detailed design. Keep the concept at a high level and focused on the core idea.
- Avoid treating the concept as a fixed solution; revisit and refine it as new evidence and feedback emerge.
- The link between the problem and the proposed solution can be weak; ensure the concept clearly responds to the challenge identified in the Scan phase.
- Teams may converge too early on one idea; explore multiple options before selecting a concept to develop further.
- The value of the solution is often unclear; clearly specify who benefits and how to strengthen alignment and testing.

Figure 3.3. Pilot concept

IMAGE OF THE SOLUTION Add a large, clear image or sketch that represents your solution.			NAME, FUNCTION & VALUE Give your pilot a name and a short description.			USE CASE Describe a common or meaningful situation where your solution will be used.																																															
KEY FEATURES & HOW THEY WORK <table border="1"> <thead> <tr> <th>Feature</th> <th>Why It Matters</th> <th>How It Works</th> </tr> </thead> <tbody> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </tbody> </table>			Feature	Why It Matters	How It Works													PHASES OR KEY ACTIVITIES <table border="1"> <thead> <tr> <th>Phase / Activity</th> <th>Purpose</th> <th>Key Actors / Tools</th> </tr> </thead> <tbody> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </tbody> </table>			Phase / Activity	Purpose	Key Actors / Tools													PUBLIC & STAKEHOLDER VALUE <table border="1"> <thead> <tr> <th>Key Actors / Tools</th> <th>Value Delivered</th> <th>Why It Matters</th> </tr> </thead> <tbody> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> <tr><td></td><td></td><td></td></tr> </tbody> </table>			Key Actors / Tools	Value Delivered	Why It Matters												
Feature	Why It Matters	How It Works																																																			
Phase / Activity	Purpose	Key Actors / Tools																																																			
Key Actors / Tools	Value Delivered	Why It Matters																																																			
TV PITCH Imagine your solution is being promoted in a 30-second TV spot.		REAL INSIGHTS & EVIDENCE Add quotes, stories, or images that inspired or informed your idea.		EXPECTED RESULTS Describe what changes or outcomes you expect over time. <table border="1"> <thead> <tr> <th>Timeframe</th> <th>What Will Change or Be Achieved</th> <th>Indicators / Examples</th> </tr> </thead> <tbody> <tr> <td>Short-term (0- 6 months)</td> <td></td> <td></td> </tr> <tr> <td>Medium-term (6-12 months)</td> <td></td> <td></td> </tr> <tr> <td>Long-term (1- 3 yeas)</td> <td></td> <td></td> </tr> </tbody> </table>				Timeframe	What Will Change or Be Achieved	Indicators / Examples	Short-term (0- 6 months)			Medium-term (6-12 months)			Long-term (1- 3 yeas)																																				
Timeframe	What Will Change or Be Achieved	Indicators / Examples																																																			
Short-term (0- 6 months)																																																					
Medium-term (6-12 months)																																																					
Long-term (1- 3 yeas)																																																					

Note: The tool canvas in full size is available in Annex A.

Source: OECD

Box 3.3. Crazy 8s: Rapid ideation tool

Based on the OECD Public Service Innovation toolkit, Crazy 8s is a quick ideation exercise that helps teams move from a defined challenge to a broad set of possible solutions. It is especially useful after reframing the problem into a “How might we...” question.

Set the focus: Write the “How might we...” question at the top of the page so everyone is responding to the same challenge.

Prepare the template: Give each participant a Crazy 8 template or an A3 sheet folded into eight sections.

Sketch eight ideas in eight minutes: Participants sketch one idea per box, spending one minute on each. Ideas should be quick, rough, and different from one another.

Avoid overthinking: Encourage participants to use simple drawings, words, arrows, or stick figures. The goal is quantity and variety, not polished design.

Share selected ideas: Each participant presents two or three ideas to the group, explaining what problem they address and why they could be useful.

Cluster similar ideas: Group related ideas together and give each cluster a short label. This helps identify common directions, promising concepts, or ideas that could be combined.

- Select ideas to develop further: Use the clusters to choose the most relevant, feasible, or interesting ideas for refinement in the next step of the pilot design process.

For more information, visit www.oecd-opsi.org

Source: OECD





Tool 4: Policy blueprint

Why it matters

The pilot concept has defined what you want to test and why it matters. The Policy Blueprint takes this concept and maps how it would actually work in full before you build anything. It is a visual model that captures the policy journey from two perspectives simultaneously: what the people your policy is designed to affect will see and experience (the frontstage), and what needs to happen inside the institution to make that possible (the backstage).

Adapted from service blueprint methodology used in service design, the Policy Blueprint is particularly relevant for policymaking pilots because it forces your team to be explicit about what you are assuming at each stage of the policy journey, and to identify which of those assumptions should be tested before committing to full implementation. This is the tool that transforms a promising concept into a structured, testable pilot design.

What it helps you do

- Map the complete journey of a policy from the perspective of those it affects.
- Connect visible user-facing interactions with the institutional processes behind them.
- Reveal dependencies across teams, units, and organisations that are not visible in a concept document.
- Surface assumptions embedded in your design at every stage.
- Identify what needs to be tested, and how, before scaling.

How to use it

1. **Define the scope.** Using the Figure 3.4, identify the policy or intervention you are mapping and the user or target group you are focusing on. A Blueprint works best when scoped to a specific user journey rather than an entire policy area. Use the outputs of your pilot Concept and challenge framing as your starting point.
2. **Map the policy journey stages.** Describe the main stages of the policy as experienced by the target group, from before they engage with it to after it has been delivered. Typical stages might include: becoming aware of the policy; accessing or applying; engaging with the main intervention; receiving the outcome; and following up. Write these across the top of your canvas as column headers, organised under Pre, During, and Post.
3. **Map the frontstage.** For each stage, capture three dimensions of what is visible to the user: what the user does (the actions they take: applying, submitting, attending, receiving, following up); what the user feels, thinks, or says (their expectations, concerns, level of trust, and , where possible, short verbatim comments); and the touchpoints they interact with (people, interfaces, systems, documents, or physical spaces). Together these three rows describe the full user experience at each stage of the journey.

4. **Map the backstage.** For each stage, identify what happens inside the institution to make the frontstage possible: the workflows, systems, databases, approval processes, and coordination mechanisms that the user does not see but that are essential to delivery.
5. **Identify actors and dependencies.** Determine which teams, units, or organisations are responsible for each backstage activity. Note where coordination across institutions is required, these are often the points where implementation breaks down.
6. **Surface assumptions.** As you map each stage, note what your team is assuming about user behaviour, motivations, capabilities, or access. For example: Do users know this policy exists? Can they access the required information? Will they trust the process? Do they have the digital literacy required? Each assumption that cannot be confirmed from existing evidence is a candidate for testing.
7. **Define testing opportunities.** For each key assumption, identify a method that could be used to test it during the pilot: interviews, surveys, usability testing, role-playing exercises, simulations, or low-fidelity prototypes. This transforms the Blueprint from a mapping exercise into a practical tool for pilot planning.

Reflection questions

- Where are the main friction points in the user journey?
- What obstacles (legal, technological, organisational) arise at each stage of the process?
- Which backstage processes create bottlenecks or coordination failures?
- Which assumptions carry the greatest risk if they turn out to be wrong?
- What should be tested first before committing to the full design?

Common challenges

- Teams tend to map the ideal version of the journey rather than the realistic one. Push for the real journey: what would actually happen at each stage given current systems and capacities?
- The backstage is often harder to map than the frontstage because it requires knowledge from across the organisation. Involve colleagues from different teams, this is a collaborative exercise, not a desk exercise.
- Do not try to map everything at once. Start with the two or three stages that carry the most uncertainty and expand from there.

Figure 3.4. Policy blueprint

	User and/or target group:	Pre Before the policy is experienced	During the policy intervention	Post After the policy has been delivered
USER JOURNEY - Visible to users experiencing the policy or the services it delivers	<p>Actions by user What the user does at each stage - applying, submitting, attending, receiving, or following up.</p> <hr/> <p>Emotional / Thinking journey: What the user feels, thinks, or says at each step - their expectations, concerns, and level of trust.</p> <hr/> <p>Touchpoints: People, interfaces, systems, documents, or spaces the user interacts with at each stage.</p>			
BACKSTAGE - Not visible to users	<p>Internal activities / tools: Workflows, approval processes, and systems that enable each stage of the journey.</p> <hr/> <p>Internal actors: Teams, units, or institutions responsible for each backstage activity, including coordination needs across organisations.</p>			
PILOT PLANNING	<p>HOW TO TEST THESE ASSUMPTIONS</p> <p>The method to test each assumption - interviews, surveys, observation, usability testing, role-play, or prototype</p>			
	<p>ASSUMPTIONS ABOUT USERS</p> <p>What the team is assuming about user behaviour, motivations, or access at each stage.</p>			

Note: The tool canvas in full size is available in Annex A.

Source: OECD



Tool 5: Testing plan

Why it matters

The policy blueprint has surfaced the assumptions embedded in your pilot design at every stage of the policy journey. Not all of them can or should be tested at once. The Testing Plan is the tool that takes those assumptions and turns the most critical ones into a structured learning agenda, defining what you will test, how, with whom, and by when, before you commit to full implementation.

The discipline of the testing plan is prioritisation: focusing your testing energy on the assumptions that are most uncertain and most consequential, rather than the ones that are easiest to address. A well-designed test is not one that confirms what you already believe, it is one that generates genuine learning, including learning that might lead you to revise your design.

What it helps you do

- Select the most relevant and impactful assumptions from your Blueprint for active testing.
- Design the simplest possible test that will generate meaningful evidence for each.
- Identify the resources and prototypes needed to carry out each test.
- Assign responsibilities and timelines so that testing actually happens.
- Generate evidence that informs each subsequent iteration of the pilot.

How to use it

1. **Identify the assumption to test** using Figure 3.5. Return to your policy blueprint and review all the assumptions you surfaced. For each row of your testing plan, select one assumption to test. Prioritise assumptions that are both highly uncertain and highly critical to the pilot's success, those where being wrong would be most costly or most difficult to correct later. Assumptions already supported by strong existing evidence, or that are low-risk if incorrect, do not need active testing at this stage.
2. **Define the activity and prototype.** For each assumption, describe the specific activity you will carry out to test it and the prototype or tool you will use. This might be a simulation of a policy process, a role-play exercise, a paper mockup of a digital interface, a drafted procedure map, or a structured interview guide. The goal is to design the simplest possible test that will generate meaningful evidence, not a polished product, but something real enough to react to.
3. **Set the objective.** Define clearly what you are expecting to gain, validate, or learn from the test. A good objective is specific: not "understand what users think" but "understand how users interact with the prototype and whether the process is clear and easy to follow." This keeps the test focused and makes it easier to analyse the findings afterwards.
4. **Identify the resources needed.** Note the human, technical, policy, or financial resources required to carry out the test. Consider: who will lead the testing, how

many hours it will require, what materials or technology are needed, and whether any institutional permissions or approvals are required before you begin.

5. **Assign deadlines and responsible persons.** Testing is a planned activity with clear accountability. For each test, define who is responsible for what and by when, including who prepares the prototype, who recruits participants, who facilitates the session, and who analyses and documents the findings.
6. **After each test: Review what you learned.** Which assumptions were confirmed? Which were challenged? What surprised you? What should change in the pilot design as a result? Use these insights to update your blueprint and pilot concept before moving to the next round of testing.

Reflection questions

- Which assumptions from the Blueprint represent the greatest risk to the success of the pilot?
- What is the simplest test that would generate meaningful learning for each?
- Are we testing with the right users and stakeholders?
- What evidence would be sufficient to lead us to change our approach?
- What can we learn that could support testing, prototyping, and future scaling?

Common challenges

- Teams are often tempted to test everything at once. Prioritise effectively: three well-designed tests on critical assumptions are worth more than ten superficial ones.
- A successful test is not one that confirms your assumptions, it is one that generates useful learning, even if that means revising your design. Build a culture within your team where disconfirming evidence is welcomed, not avoided.
- Do not wait for a perfect prototype before testing. The purpose of early testing is to learn, not to impress.



Figure 3.5. Testing Plan

<p>ASSUMPTION TO TEST</p> <p>Main assumptions to start our testing. Example: Simulation of service delivery at the citizen service counter to observe user interaction with forms, waiting times, and guidance received.</p>	<p>ACTIVITY AND PROTOTYPE</p> <p>Prototype and activities you will carry to test the assumption Example: Simulation of service delivery at the citizen service counter to observe user interaction with forms, waiting times, and guidance received.</p>
<p>OBJECTIVE</p> <p>What are you expecting to gain, validate or learn? Example: Understand how people interact with the prototype and whether it is clear and easy to use.</p>	
<p>RESOURCES</p> <p>What human, tech, policy or monetary resources are needed for the test Example: Who will lead the testing, how many hours it will require, materials or tech we need for the test.</p>	<p>DEADLINES / RESPONSIBLE PERSONS</p> <p>Who will be responsible of what and when. Example: Define responsible persons, deadlines, and next steps.</p>

Note: The tool canvas in full size is available in Annex A.
Source: OECD



Tool 6: Low-fidelity prototyping

Why it matters

Your testing plan has defined what you need to learn and identified the prototypes you will use to learn it. Now you need to build them. A prototype is a simplified, testable version of your policy idea that makes an abstract concept tangible, something that can be shown to users, reacted to, and improved. The testing plan told you what to test; this tool tells you how to build what you need to test it with.

Low-fidelity prototypes are deliberately rough. They are not finished products, they are thinking tools. Their purpose is to generate feedback quickly and cheaply, before you have invested significant resources in a particular direction. In a public administration context, this might mean a printed mockup of a new form or process, a hand-drawn sketch of a digital interface, a drafted procedure map, a one-page scenario description, or a role-play script simulating a new way of working. The format matters less than the principle: build something real enough to react to before committing to the final design.

What it helps you do

- Transform an abstract policy concept into something that can be seen, discussed, and reacted to.
- Surface problems with your design early, when they are still cheap to fix.
- Generate specific, actionable feedback rather than general impressions.
- Communicate your concept to stakeholders more effectively than written descriptions alone.
- Produce the evidence you committed to gathering in your testing plan.

How to use it

1. **Choose your fidelity level** using Figure 3.6. A first prototype should be low-fidelity: a hand-drawn sketch, a printed mockup, a simple process flowchart, or a one-page scenario description. The goal is speed and feedback, not polish. Resist the temptation to invest in a high-quality version before you have tested the underlying concept, a rough prototype generates more honest reactions than a polished one, because users respond to what they see rather than to what they think you want to hear.
2. **Choose your prototype format.** Different assumptions call for different formats. Refer back to your testing plan to match the format to the assumption you are testing. To test whether users understand a new process, a step-by-step flowchart or a role-play scenario works well. To test whether a digital interface is intuitive, a paper sketch or a simple clickable mockup is sufficient. To test whether a new procedure is operationally feasible, a process map with internal actors and workflows is more useful than a visual design.
3. **Build the prototype.** Keep it simple and build it quickly, a prototype that takes weeks to produce defeats the purpose. Use whatever materials are available: paper, sticky notes, a basic slide deck, or a simple word-processed document.

If the prototype involves a process rather than a product, write a short scenario script that describes what happens at each step and assign roles to team members for a role-play exercise. If you are working on a policy intervention or strategic document rather than a digital tool or service, the examples in Box 3.3 illustrate what low-fidelity prototyping can look like in that context.

4. **Prepare for the test.** Brief the team on what you are testing and what you are looking for. Identify five to eight users from your target group, these should be real users, not colleagues who are already familiar with the concept. Decide in advance who will facilitate, who will observe and take notes, and what you will do if users get stuck.
5. **Run the test.** Keep the test short and focused: unlike later pilot implementation or evaluation activities, low-fidelity prototype testing should usually take only enough time to observe whether users understand the concept, where they get confused, and what needs to change. Show the prototype to users without explaining it first. Let them try to understand it independently. Observe where they get confused, what they do naturally, and what questions they ask. Resist the urge to help or explain; their confusion is your most valuable data.
6. **Collect structured feedback.** After the observation, ask: what worked? What was confusing? What would you change? What is missing? Take notes on both what users say and what you observed them doing, these are often different, and the gap between them is frequently the most instructive finding.
7. **Revise and iterate.** Update your prototype based on the feedback and test again. Expect to go through at least two rounds of iteration before moving to higher-fidelity versions. Each round should produce a clearer, more refined concept, and each round of testing should focus on the assumptions that the previous round left unresolved.

Reflection questions

- Are we testing with real users from the target group, or with colleagues who already understand the concept?
- What did users do differently from what we expected, and what does that tell us about our design?
- Which assumptions did this round of testing confirm, and which did it challenge?
- What is the one change we should make before testing again?

Common challenges

- Teams are often reluctant to show unfinished work in formal institutional settings. Reframe this explicitly with your team and your stakeholders: the purpose of a prototype is to generate feedback, and a rough version produces more honest reactions than a finished one.
- Testing with the wrong people, senior managers rather than end users, for example, produces feedback that reflects authority rather than experience. Be deliberate about who you test with, and refer back to the Scan phase if you need to identify the right participants.

- Do not prototype everything. Focus on the elements of your design that carry the most uncertainty, as identified in your testing plan. Prototyping details that are already well understood is a use of time that could be better spent elsewhere.
- Iteration requires discipline. After each round of testing, make a decision: what changes, and what stays. Avoid the temptation to keep revising without testing, the only way to know if a change has improved the design is to test it with users again.

Figure 3.6. Low-fidelity prototyping

WHAT TO BUILD	WHAT TO TEST	WHO TO TEST WITH	WHAT YOU LEARNED
<p>Format: The type of prototype you will build — sketch, flowchart, scenario, role-play script, or slide deck.</p> <p>Fidelity: Keep it as rough as possible. The rougher it is, the more honest the feedback.</p>	<p>Assumption: The specific assumption from your Testing Plan this prototype is designed to test.</p> <p>Learning objective: What a successful test would tell you — in one sentence.</p>	<p>Participants: Five to eight real users from your target group. Not colleagues already familiar with the concept.</p> <p>Roles: Who facilitates, who observes, who takes notes.</p>	<p>What worked: What users responded to positively or navigated without difficulty.</p> <p>What was confusing: Where users got stuck, asked questions, or made unexpected choices.</p> <p>What to change: The one adjustment to make before testing again.</p> <p>Next step: Revise and retest, or move forward.</p>

Note: The tool canvas in full size is available in Annex A.
Source: OECD

Box 3.3. What does prototyping look like for a policy or strategic document?

Prototyping is not only for digital tools or service interfaces. Policy interventions and strategic documents can be prototyped too, and the earlier you do it, the cheaper it is to discover what is not working. The core principle is the same: build the simplest possible version of your idea that can be shown to real users and generate honest feedback. In a policy context, this might look like any of the following:

- **A one-page policy summary.** Draft the core logic of your proposed intervention on a single page, the problem, the proposed response, who it affects, and what it requires, and share it with a small group of affected users or implementers. Ask: is this clear? Does it reflect your experience of the problem? What would prevent this from working in practice?
- **A process map or flowchart.** Draw out how the policy would work step by step, what a civil servant or citizen would need to do at each stage, and what the institution would need to do behind the scenes. This is a lightweight version of the policy blueprint and is particularly useful for testing whether a proposed procedure is operationally realistic.
- **A role-play simulation.** Script a scenario based on your proposed policy and act it out with team members or stakeholders playing the roles of the people it would affect. This is especially effective for testing new regulations, procedures, or cross-institutional workflows where the interactions between actors matter as much as the content of the policy itself.
- **A storyboard.** Create a simple visual sequence, even hand-drawn, showing what a person's experience would look like before and after the policy is in place. This is useful for building shared understanding within your team and for communicating the concept to stakeholders who find written policy documents hard to engage with.
- **A tabletop simulation.** Bring together a small group of implementers and work through a hypothetical scenario under the proposed policy rules. What decisions would need to be made? Where does the logic break down? What information is missing?
- **A simple working prototype.** In some cases, the prototype may be a basic functional version of the solution, such as a simple feature developed within an IT tool, a draft digital form, a clickable mock-up, or a limited version of a platform. The aim is not to build the final product, but to make one core function tangible enough for users to test and react to.

These approaches are used worldwide in innovative policymaking pilots. None of them needs to be perfect on the first attempt; prototypes are designed to evolve. A one-page summary can become a more detailed process map after the first round of feedback; a role-play scenario can become a structured pilot procedure after testing reveals what works. Each iteration can increase in fidelity as confidence in the design grows, moving from rough sketches to more developed versions only once the underlying logic has been validated with real users. For example, in Colombia, Bogotá's Secretariat of Women's Affairs tested the Care Blocks concept in participatory sessions with caregivers before any infrastructure was built, refining the model through successive rounds of feedback before committing to physical locations and resource allocation (see more in Box 3.7) (OECD, 2023).

Source: OECD.



Tool 7: Peer exchange and feedback

Why it matters

You have built and tested a prototype and refined your concept through user feedback. Now it is time to expose it to a different kind of scrutiny: the perspectives of peers working on similar challenges in other institutions or contexts, who can see what your team, through familiarity with the problem, may no longer be able to see.

Teams running pilots can become attached to their own ideas and overlook weaknesses that are visible to anyone approaching the concept fresh. The Ritual Dissent technique provides a structured format for generating genuine critical feedback rather than polite agreement, helping teams identify blind spots, implementation barriers, and opportunities for improvement before moving into the Scale phase. The purpose is not to defend the idea but to learn from diverse perspectives and strengthen the solution through constructive dissent. Pilot team can use this activity across the whole Scan-Pilot-Scale process.

What it helps you do

- Surface blind spots and weaknesses that internal review consistently misses.
- Stress-test your implementation and scaling logic before committing to full deployment.
- Access knowledge and experience not available within your own team or institution.
- Strengthen your pilot design through constructive challenge rather than confirmation.

How to use it

1. **Gather the materials** you will use for your presentation. Unlike the other tools in this guide, this activity does not use a dedicated canvas. Teams can use the tools' canvases previously developed or a short slide deck.
2. **Present the pilot.** One team member briefly presents the pilot concept, the prototype, the implementation strategy, and the scaling ambitions. Keep the presentation short, no more than ten minutes. The goal is to give peers enough context to meaningfully challenge the design, not to convince them it is correct.
3. **Listen to critical feedback.** The presenting team remains mostly silent while peers discuss the proposal, challenge assumptions, and identify risks or gaps. Assign one team member to take detailed notes on all feedback, including comments that feel uncomfortable or challenge the core concept. The discipline of not defending the idea in real time is what makes this technique work: it forces the presenting team to listen rather than explain, and allows peers to speak more candidly than they would in a conventional review session.

4. **Reflect and improve.** After the session, the team reviews the feedback together and identifies which insights should inform the next iteration of the pilot. Not all feedback will be equally relevant or feasible to act on, but engaging with it seriously, rather than dismissing it, is what produces a stronger design. For each insight you decide to act on, note what will change. For each insight you decide not to act on, note why.
5. **Iterate the necessary outputs.** Using the feedback from this session, teams can refine their outputs, including testing plans and prototypes as many times as needed.

Common challenges

- Sessions can drift into general discussion if not well facilitated. Share focused questions with participants in advance and assign a facilitator to keep the conversation anchored to the specific aspects of the design you want challenged.
- Teams can become defensive when their design is challenged. Brief your team beforehand: a challenge that improves your design before implementation is far less costly than a problem that emerges after it.
- When peers share examples from other contexts, always ask: what would need to be different for this to work in the Bulgarian public administration context?





Tool 8: Theory of change

Why it matters

Your pilot has been designed, tested through prototyping, and strengthened through peer feedback. Before moving into the Scale phase, you need to articulate clearly, in a form that others can understand and scrutinise, how and why you believe this pilot will create the change you are aiming for. That is what the theory of change does.

A theory of change is a structured account of the logic connecting your inputs and activities to the outputs, outcomes, and long-term impact you are aiming for. It makes the causal reasoning behind your pilot explicit: if we do this, and these conditions hold, then this change will follow. It is not merely a planning exercise. It is a tool for testing the internal coherence of your pilot design, communicating it clearly to decision-makers and partners, building support for the innovation being tested, and defining what you will measure when you move into the Scale phase. By making the expected change visible and understandable, a well-developed theory of change can also support early change management, helping stakeholders see their role in the pilot and increasing their willingness to engage with it. A strong theory of change ensures that your pilot is not only innovative but also evidence-informed, impact-oriented, and easier to mobilise around.

What it helps you do

- Articulate the causal logic of your pilot in a form that others can understand and challenge.
- Identify assumptions embedded in your causal chain that still need to be validated.
- Align team members and stakeholders around a shared understanding of what success looks like.
- Define the indicators that your evaluation framework will measure in the Scale phase.
- Build a stronger case for institutional adoption and scaling.

How to use it

1. **Define the project purpose** using Figure 3.7. Start by describing the overarching change the pilot seeks to achieve, the ultimate objective and the public value it will generate. Be specific and realistic about scope: a pilot on cross-generational mentoring in one agency might meaningfully reduce the loss of institutional knowledge in participating organisations and provide a replicable model for others, but it will not transform the entire public administration.
2. **Identify the problems and needs.** Drawing directly from your Problem Framing tool, describe the problem the pilot seeks to address. Who is affected? What unmet needs exist? What barriers currently prevent better outcomes? Every problem identified here should connect to an intended outcome later in the chain.
3. **Map inputs and activities.** Identify the resources and actions required to deliver the pilot. Inputs may include staff and expertise, technology and infras-

structure, funding, partnerships, and data. Activities are the actions undertaken using these resources, training, outreach, prototyping, stakeholder engagement, or policy implementation.

4. **Describe the products and services.** Identify the tangible outputs produced through the pilot that users and stakeholders will directly experience, a digital platform, a new process, training materials, guidance documents, a competency framework, or a set of implementation tools.
5. **Define intermediate results.** Describe the short- and medium-term changes that should occur if the pilot works as intended, the first signs that it is creating value. Examples include increased use of a service, improved skills among participants, changes in behaviour, better coordination among stakeholders, or greater trust or engagement.
6. **Define final results.** Describe the longer-term changes the pilot seeks to contribute to, improved public service performance, reduced administrative burden, increased institutional capacity, or stronger public trust. Connect these back to the problems and purpose defined at the start of the chain.
7. **Surface assumptions and risks.** Review the connections between each stage of the chain and identify what needs to be true for each link to hold. Many of these assumptions will already be familiar from your Policy Blueprint and Testing Plan. Also note risks, factors outside your control that could disrupt the causal pathway and prevent the desired results from being achieved.
8. **Design monitoring indicators.** For each of the upper rows, inputs, activities, and products, identify at least one indicator that tracks whether implementation is proceeding as planned. Examples include the number of users engaged, sessions delivered, or the percentage of staff trained. Where possible, complement these with early outcome-oriented indicators that capture whether the pilot is beginning to create value for users or the organisation, such as user satisfaction, reduced processing time, improved confidence, or fewer errors. Box 3.4 illustrates what well-designed monitoring and outcome indicators look like in practice.
9. **Design outcome indicators.** For each of the lower rows, intermediate and final results, identify at least one indicator that measures whether the pilot is achieving its intended changes. Examples include user satisfaction levels, behavioural change, reduced processing times, or increased service uptake. These indicators become the foundation of your evaluation framework in Tool 9 in the Scale phase.

Reflection questions

- Is there a clear and plausible connection between each link in the causal chain?
- Which assumptions are most critical to the success of the pilot, and have they been tested?
- Are we measuring both what we are doing and what change we are creating?
- How will this theory of change support future decisions about scaling?

Common challenges

- Theories of change can become over-complicated when multiple stakeholders contribute to the logic. Keep the core chain on one page and move supporting detail to an accompanying narrative.
- Avoid confusing outputs with outcomes. An output is what your pilot produces, a training programme, a platform, a set of guidelines. An outcome is the change that results from people using that output, increased skills, changed behaviour, improved decision-making. Both matter, but they measure different things and require different indicators.
- The theory of change should be treated as a live document. If the pilot produces evidence that challenges one of the causal links, update it. A fixed theory of change that is never revised is not being used as a learning tool.
- Teams sometimes develop a theory of change at the end of the Pilot phase as a retrospective description of what they did, rather than a prospective account of what they expect to happen. The tool is most valuable when completed before implementation, not after.

Box 3.4. What do good indicators look like?

Monitoring indicators and outcome indicators measure different things and should be designed differently.

Monitoring indicators track whether implementation is proceeding as planned. Good monitoring indicators are:

- Countable and time-bound, they can be measured within the pilot timeframe.
- Linked to specific activities or outputs, not to changes in the world beyond your control.
- Simple enough to collect without a dedicated evaluation team.

Examples: number of users engaged in testing sessions; number of sessions delivered; percentage of staff trained; number of institutions participating.

Outcome indicators measure whether the pilot is creating the intended change. Good outcome indicators are:

- Linked to a specific change in behaviour, capacity, satisfaction, or performance.
- Measurable before and after the pilot so that change can be assessed.
- Realistic for the scope and timeframe of the pilot, a small pilot cannot be expected to produce system-wide change.

In practice, the distinction matters because they require different data-collection approaches and timelines. Bogota's Care Blocks programme illustrated this clearly. Monitoring indicators tracked the number of Care Blocks opened, and the number of caregivers reached each year, while outcome indicators measured reductions in time poverty among participating women and increases in their access to education and wellbeing services, changes that required longer timeframes and a different data collection approach to capture (see more in Box 3.7) (OECD, 2023).

Source: OECD.

Figure 3.7. Theory of change

	PROJECT PURPOSE	MONITORING INDICATORS
	PROBLEMS / NEEDS	
↓		
	INPUTS / ACTIVITIES	
↓		
	PRODUCTS / SERVICES	OUTCOME INDICATORS
↓		
	INTERMEDIATE RESULTS	
↓		
	FINAL RESULTS	
	ASSUMPTIONS THAT UNDERPIN THE ABOVE	

Note: The tool canvas in full size is available in Annex A.
Source: OECD

Box 3.5. What does the Pilot phase look like in practice?

BULGARIA: MAPPING ASSUMPTIONS AND BUILDING A TESTABLE PROTOTYPE - INSTITUTE OF PUBLIC ADMINISTRATION AND PATHWISEAI

As part of the pilots run through the Strengthening the Capacity to Design and Monitor Public Policies project, the Institute of Public Administration piloted PathwiseAI, an initiative to provide civil servants with personalised, competency-based learning pathways.

The team used a blueprint to map the full journey of a civil servant from identifying a training need to completing a learning pathway. The mapping supported surfacing a critical backstage dependency: the proposed platform required integration with the IPA's Moodle-based e-learning infrastructure, which was itself undergoing an upgrade. This constraint, which was invisible in the original concept, reshaped the pilot's scope and timeline before anything was built.

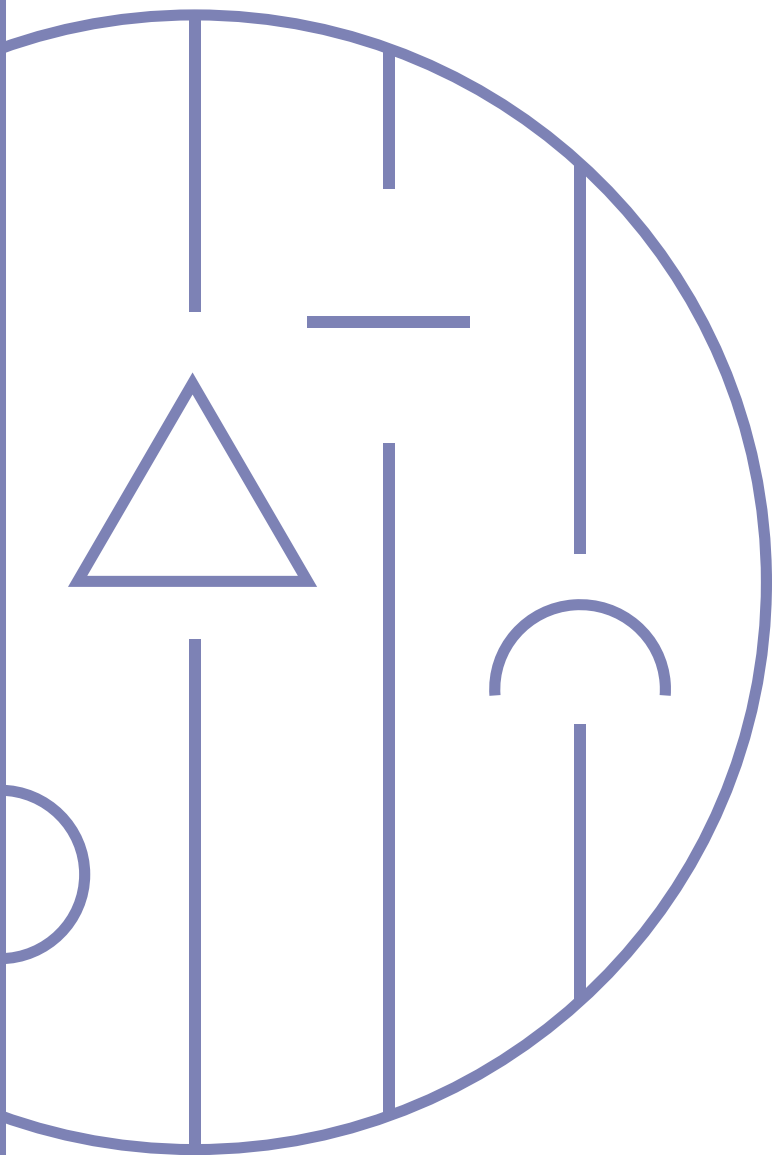
The blueprint also surfaced assumptions that had not yet been tested: that civil servants would voluntarily engage with a competency self-assessment; that managers would support participation; and that cybersecurity was a representative first competency area for trial. Each was built into a structured testing plan and user questionnaire. The Pilot phase produced a conceptual prototype, a service blueprint, a testing plan, and a competency framework example, transforming an ambitious concept into a structured, testable design grounded in institutional reality (OECD, 2026).

INTERNATIONAL EXAMPLE: ITALY'S INPS AI EMAIL SORTING TOOL

When Italy's National Institute of Social Security (INPS) faced a surge in citizen emails during the COVID-19 pandemic, it began by mapping the full email journey across 448 territorial offices, from the moment a citizen sent a request to the moment an employee responded. The mapping revealed that the bottleneck was not in responding to emails but in classifying and routing them: a backstage workflow problem, not a communication one.

Before building, the team tested critical assumptions: that the AI could achieve sufficient accuracy; that employees would trust and adopt it; and that volume gains would offset transition costs. The tool was deployed gradually, first in a small number of cities, then expanded, with employees at the centre throughout. More than two-thirds of emails are now automatically dispatched with 80% accuracy, freeing employees to focus on complex cases requiring human judgement (OECD, 2024).

Source: OECD.



SCALE PHASE



SCALE PHASE

The Scale phase is where you move from a tested pilot to a plan for sustained, broader implementation. The question is no longer whether your idea works in a controlled setting; the Pilot phase has generated evidence on that. The question now is: how do we know it created sufficient value to justify scaling? What conditions need to be in place for it to work beyond the pilot? And how do we make it last?

Scaling is not simply doing more of the same. Solutions that work in one organisational context often need adaptation to work in another. Resources, partnerships, and governance structures that were informal during the pilot need to become explicit and durable. And the evidence generated during the pilot needs to be translated into a compelling case for decision-makers who were not part of the process. Box 3.7 illustrates how the Scale phase was applied in practice in Bulgaria and beyond.

MAIN PURPOSE:

Plan for broader implementation and sustained impact. By the end of this phase, you should have a clear evaluation framework that captures what your pilot achieved and why, a structured scaling strategy that identifies what needs to change for the solution to work at full scope, and a credible implementation model that can be presented to decision-makers, partners, and funders.

WHAT ARE THE KEY BENEFITS OF APPLYING THESE TOOLS?

- **Tool 9 - Evaluation framework:** Defines what success looks like in measurable terms and produces the evidence needed to justify scaling, moving from anecdotal impressions to a credible, structured account of what the pilot achieved and why.
- **Tool 10 - Scaling plan:** Translates the lessons of the pilot into a concrete roadmap for broader implementation, identifying what is transferable, what needs to adapt, and what enabling conditions need to be created for the solution to work at scale.

Tool 9: Pilot evaluation framework

Why it matters

Your theory of change has defined the causal logic of your pilot and identified the indicators that should tell you whether change is occurring. The Pilot evaluation framework is the tool that puts that logic into practice, defining how you will collect evidence, at what points, and against what baseline, so that by the end of the pilot you have a structured, credible account of what worked, what did not, and what would need to change to improve results at scale.

Evaluation frameworks for pilots differ from traditional programme evaluations in one important way: they are not primarily about accountability, proving that the pilot achieved its targets, but about learning, understanding why it did or did not work and what that means for the next stage. A good evaluation framework produces evidence that is useful for decision-making, not just for reporting.

What it helps you do

- Define what success looks like for your pilot in terms that can be measured.
- Establish baselines before implementation so that change can be assessed.
- Design data collection methods that are proportionate to the scale of the pilot.
- Produce evidence that decision-makers need to decide whether to scale.
- Create a foundation for continuous improvement beyond the pilot.

How to use it

1. **Define your evaluation questions.** Identify the two or three most important things you want to know at the end of the pilot, making sure they are aligned with the pilot's original objectives. These questions should link directly to the assumptions in your theory of change and reflect what decision-makers most need to know to decide whether to scale. Good evaluation questions are specific: not "did the pilot work?" but "did the pilot reduce the time civil servants spent on repetitive procedural queries, and by how much?"
2. **Select your indicators.** Drawing directly from the indicators you defined in your theory of change, select the monitoring indicators that will track implementation and the outcome indicators that will measure change. Refer to Box 3.4 for guidance on what good indicators look like. Aim for three to five well-chosen indicators rather than a long list measured superficially.
3. **Establish baselines.** For each indicator, identify what the current state is before the pilot begins. Without a baseline, you cannot assess whether change has occurred. Baselines can often be drawn from existing administrative data, previous surveys in public administration and beyond, or user research conducted during the Scan phase.
4. **Design your data collection approach.** Identify how you will gather data for each indicator, at what points during and after the pilot, and who is responsible. A simple data collection plan, listing each indicator, the method, the

frequency, and the responsible person, keeps the evaluation manageable and ensures it actually happens.

5. **Define decision thresholds.** Identify what level of evidence would be sufficient to recommend scaling. Being explicit about this in advance, before you have seen the results, prevents post-hoc rationalisation and strengthens the credibility of your recommendation. For example: if user satisfaction with the new process is above 70% and processing time is reduced by at least 20%, we will recommend scaling to two additional public organisations.
6. **Plan for learning, not just reporting.** Build in at least one mid-point review during the pilot where you assess emerging findings and use them to refine the implementation. Evidence collected only at the end of a pilot is useful for accountability but not for improvement. Regular review points make the evaluation a learning tool rather than a compliance exercise.

Reflection questions

- Are our evaluation questions specific enough to produce actionable findings?
- Do we have baseline data for each indicator, and if not, how will we establish it before implementation begins?
- Are we measuring both what we are doing and what change we are creating?
- What would the evidence need to show for us to recommend scaling confidently?

Common challenges

- Evaluation frameworks are often designed too late, after the pilot has started, making baseline data collection limited. Build the framework before implementation begins, using the indicators already defined in your theory of change.
- Avoid the temptation to measure everything. Three to five well-chosen indicators, measured rigorously, are more useful than twenty indicators measured weakly.
- Ensure data collection responsibilities are clear. Assign a named person to each indicator and build data collection into regular workflows rather than treating it as an additional task.
- Do not treat the absence of problems as proof of success. Ensure your evaluation captures positive change, not just the lack of failure.

Figure 3.8. Pilot evaluation framework

EVALUATION QUESTIONS		INDICATORS	
<p>What are the two or three most important things you want to know at the end of the pilot? Link each question directly to the assumptions in your theory of change.</p>		<p>Monitoring indicators: What will you count or measure to track whether implementation is proceeding as planned? Link each indicator to a specific activity or output.</p> <p>Outcome indicators: What will you measure to assess whether the pilot is creating the intended change? Link each indicator to a specific intermediate or final outcome from your theory of change.</p>	
BASELINES			
<p>What is the current state of each indicator before the pilot begins? Where will you find this data, administrative records, previous surveys, or user research from the Scan phase?</p>			
DATA COLLECTION			
<p>Method: How will you collect data for each indicator, administrative records, surveys, interviews, observation, or service data?</p>	<p>Frequency: When will you collect data, before the pilot begins, at a mid-point review, and at the end?</p>	<p>Responsible person: Who is accountable for collecting and recording the data for each indicator?</p>	
DECISION THRESHOLDS			
<p>What level of evidence would be sufficient to recommend scaling? Define this before you see the results.</p>			

Note: The tool canvas in full size is available in Annex A.
Source: OECD

Tool 10: Scaling funnel

Why it matters

One of the most common reasons innovative pilots remain pilots, tested once, documented, and then quietly set aside, is that scaling was never planned for during the pilot itself. By the time the pilot concludes, the team may have moved on, the budget may have been spent, and the institutional moment may have passed. The conditions that made the pilot possible, a committed leader, a protected team, specific skills, and a particular partnership, were not necessarily designed to last beyond the original timeframe. Scaling may also require different resources, competencies, governance arrangements, or even a different team from the one that ran the pilot. Without anticipating these changes, no one builds the case for what should come next..

This is why the Scaling Plan is most valuable when prepared during the pilot, not after. Thinking about scale early changes how you design the pilot: it pushes you to document what you are doing in ways that others can replicate, to test whether the solution works beyond the specific conditions of the pilot context, and to identify the enabling conditions that would need to exist in other institutions before they could adopt it. A pilot designed with scaling in mind is fundamentally different from one designed solely to answer whether an idea works.

The Scaling Funnel canvas provides the practical framework for this planning. It asks you to describe how your solution could realistically grow across four time horizons: today, one year from now, two years from now, and five years from now, across five dimensions: reach, institutional adoption, maturity, resources and enablers, and impact. Start with today's pilot and imagine realistic growth over time, using numbers wherever possible.

What it helps you do

- Define a realistic pathway from pilot to broader implementation across four time horizons.
- Identify what is transferable from the pilot and what needs to adapt for different contexts.
- Map the resources, partnerships, and governance structures required at each stage of growth.
- Communicate a credible scaling roadmap to decision-makers, partners, and funders.
- Identify risks that could prevent scaling and propose mitigation strategies at each stage.

How to use it

1. **For each time horizon on the Figure 3.9:** today, one year from now, two years from now, and five years from now, complete the five dimensions below. Start with an honest description of today's pilot before projecting forward.
2. **Define the reach.** Describe how many people benefit from the solution at each stage, the number of users, citizens, civil servants, regions, or institutions

covered. Start with the current pilot scope and project realistic growth. For example, today: 200 users in one department; one year from now: 3,000 users across three agencies; two years from now: 50,000 users with national rollout to regional offices; five years from now: public administration-wide coverage.

3. **Map institutional adoption.** Identify which departments, agencies, municipalities, or ministries are using or implementing the solution at each stage. Ask: who needs to adopt this for it to reach the next level of scale? What would it take to bring them on board? For example, today: prototype tested in one department; one year from now: three agencies adopt the solution; two years from now: national rollout to regional offices; five years from now: standard across government.
4. **Assess the maturity of the solution.** Describe the stage of development of the solution at each time horizon, from prototype through to institutionalised policy or practice. Being honest about current maturity helps identify what still needs to be developed before the next stage of scaling is viable. For example, today: prototype tested; one year from now: operational service; two years from now: integrated into official processes; five years from now: institutionalised policy or programme.
5. **Identify resources and enablers.** For each time horizon, identify the budget, technology, partnerships, legal frameworks, political support, and competencies required to sustain and grow the solution. Ask: what is already in place from the pilot, and what additional resources or conditions are needed for the next stage? Who needs to be involved beyond the original pilot team, and what level of engagement is required from other organisations that will adopt, support, or be affected by the solution? For example, today: innovation team and pilot funding; one year from now: dedicated implementation team and active engagement from partner organisations; two years from now: budget line, legal framework, and trained teams in adopting institutions; five years from now: permanent funding, governance structure, and cross-government ownership.
6. **Define the impact.** Describe the measurable improvement in service, efficiency, access, or outcomes expected at each stage. Focus on concrete, realistic changes that connect directly to the outcome indicators defined in your Evaluation Framework. For example, today: proof of concept; one year from now: 50% faster service delivery; two years from now: improved access across regions; five years from now: structural improvement in service delivery.
7. **Identify what needs to change between stages.** Review the gap between each time horizon and identify the regulatory, resource, organisational, or partnership conditions that need to be in place before the next stage is possible. Which of these are within your team's control, and which require action by others, senior leadership, other institutions, or external partners?
8. **Conduct a risk analysis.** For each stage of the funnel, identify the main risks that could prevent progress, political, financial, technical, regulatory, or operational, and propose mitigation strategies. Consider risks to the scaling process itself and risks to the sustainability of the solution once it has been scaled.

9. **Build the case for scaling.** Draw on your evaluation evidence to construct a clear argument for moving forward. A good scaling case acknowledges what the pilot demonstrated, what it did not demonstrate, and what would need to be true for each subsequent stage to succeed. Decision-makers are more persuaded by honest accounts of evidence and remaining uncertainty than by overconfident claims. With your scaling plan in place, the next step is communicating it effectively to the people whose support you need to move forward. Box 3.6 offers practical guidance on how to do this.

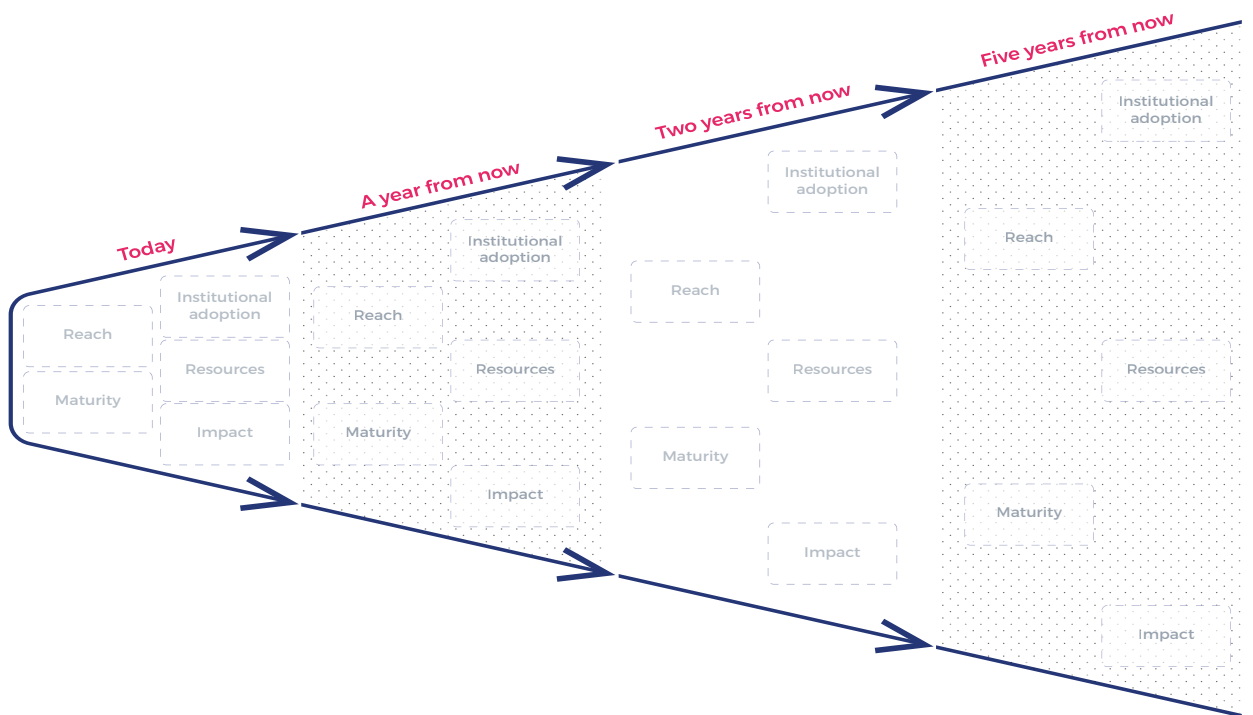
Reflection questions

- Did we start thinking about scaling early enough, or are we retrofitting a scaling plan onto a pilot that was not designed with scaling in mind?
- Is the growth projected in the funnel realistic given current political, financial, and operational conditions?
- Are the conditions that enabled the pilot's success replicable at scale, or were they specific to this pilot context?
- What would need to be institutionalised, in governance, budget, or regulation, for the solution to survive a change in leadership or political priorities?
- What is the minimum viable next step, the smallest move beyond the pilot that would generate meaningful additional evidence and keep momentum alive?

Common challenges

- The most common reason pilots remain pilots is that no one planned for scaling until it was too late. If you are completing this tool after the pilot has concluded, use it to reconstruct the scaling logic retrospectively, but commit to starting earlier next time.
- The funnel can invite overoptimistic projections. Keep each time horizon grounded in realistic assumptions about resources, political conditions, and institutional capacity. An honest funnel that identifies genuine constraints is more useful than an ambitious one that cannot be delivered.
- Scaling plans sometimes assume that the conditions that made the pilot successful will persist at scale without deliberate effort. Leadership changes, budget cycles end, and motivated teams move on. These conditions need to be explicitly designed for and institutionalised at each stage of the funnel, not assumed.
- Cross-institutional scaling requires partnership agreements and governance arrangements that take longer to negotiate than anticipated. Start those conversations during the pilot, not after it, and build negotiation time into your roadmap.
- Be honest about what the pilot did not demonstrate. The strongest argument for scaling is a concrete one, here is what we learned, here is what we still do not know, and here is why the evidence is sufficient to justify the next step.

Figure 3.9. Scaling funnel



Note: The tool canvas in full size is available in Annex A.
Source: OECD

Box 3.6. Communicating your pilot: every conversation is an opportunity

Innovative policymaking pilots need support, from users, public servants, leaders, and partners, at every stage of the Scan-Pilot-Scale journey. Building a clear, compelling story about your pilot is not something you do once at the end of the process for a formal presentation. Every conversation with a decision-maker, every team meeting with a potential partner, and every informal exchange with a senior leader or other public servants is an opportunity to advance understanding and build buy-in. The teams whose pilots gain institutional support are usually the ones who can explain what they are doing, why it matters, and what they need, clearly and concisely, at any moment, with or without a slide deck.

This means it is always worth having a clear pitch in your head, ready to use whenever the opportunity arises.

WHAT A TYPICAL PILOT PITCH INCLUDES:

- The problem or need. What challenge does the pilot address, and why does it matter now? Ground this in evidence, a survey finding, a user insight, a strategic priority, rather than a general statement about what needs to improve.
- The people affected. Who experiences the problem most directly? Making the human dimension of the challenge visible is what turns an abstract policy issue into something a decision-maker can connect with.
- The proposed solution. What are you testing, and how does it address the problem? Keep this concise, one or two sentences describing what the pilot does and how it works.

- What was tested or learned. What evidence have you generated so far? What did you find out, and how has it shaped the design? This is what distinguishes a pilot from an idea, the fact that you have already done something and learned from it.
- The expected public value. What will change if the pilot succeeds? For whom, and by how much? Be specific and realistic rather than ambitious and vague.
- The support needed to move forward. What are you asking for, a decision, a resource, an introduction, a signal of support? Every pitch should end with a clear and specific ask.

BEST PRACTICES FOR COMMUNICATING YOUR PILOT:

- Know your audience. A pitch to a minister differs from a conversation with a peer team or a potential implementing partner. Consider what matters most to the person you are talking to and frame your story accordingly, if possible.
- Lead with the problem, not the solution. Decision-makers are more persuaded by a clearly articulated challenge than by a detailed description of a proposed fix. Make them feel the problem before you offer the answer.
- Use concrete examples. A specific user story, a striking finding from your testing, or a short quote from a pilot participant is more compelling than a process diagram or a framework slide.
- Be honest about uncertainty. Acknowledging what you do not yet know builds credibility. A pitch that claims certainty about untested assumptions will be challenged, and rightly so. The strongest pitches combine confidence in what has been learned with honesty about what remains to be validated.
- Keep it short. If you cannot explain your pilot in two minutes, you do not yet understand it well enough yourself. Practice until you can.
- Prepare for questions. The questions a decision-maker asks after a pitch are often more important than the pitch itself. Think through the three or four challenges most likely to arise and have clear, honest answers ready.

For teams that want to develop their communication further, the OECD Public Service Innovation Booster includes a dedicated Storytelling tool that supports teams in building a complete narrative of their process, from the original scope and research insights to the proposed concept and the specific ask, and in tailoring that narrative for different audiences. The tool is available at oecd-opsi.org.

Source: OECD.

As teams move from planning to implementation, the tools in this guide provide a strong foundation, but the work of scaling does not end with a roadmap. For teams ready to take the next step and describe, in concrete operational terms, how their solution will function once implemented, the OECD Toolkit Navigator offers additional tools for outlining the implemented policy model, such as the Model of the Implemented Solution or the Policy Model Canvas. These tools help teams define the full operational architecture of their solution. These tools are available at oecd-opsi.org.

Box 3.7. What does the Scale phase look like in practice?

BULGARIA: DESIGNING FOR SCALE FROM THE OUTSET - NATIONAL EMPLOYMENT AGENCY AND THE MENTORING PROGRAMME

As part of the pilots run through the Strengthening the Capacity to Design and Monitor Public Policies project, the National Employment Agency developed a cross-generational mentoring programme to address the loss of institutional knowledge as experienced civil servants approach retirement.

What distinguished this pilot in the Scale phase was that scaling had been built into the design from the outset rather than added as an afterthought once the pilot was deemed successful. The programme structure, tools, and implementation guides were all developed to be adaptable to other institutions. The Scale phase explicitly addressed three questions: which elements are transferable as-is (the core mentor-mentee matching framework, the training curriculum, and the evaluation tools); which need institutional adaptation (the onboarding process, the frequency of sessions, and the communication strategy); and what enabling conditions are required at any adopting institution (senior leadership endorsement, protected time for participation, and a designated programme coordinator).

A live Masterclass for Mentors, delivered to civil servants from across the country, including participants from the Institute of Public Administration and the Ministry of Foreign Affairs, served as a dissemination prototype: a real-world test of whether the programme could be transferred beyond the National Employment Agency to partner organisations. The cross-institutional ambition has made scaling a more plausible continuation of the pilot process rather than a separate challenge (OECD, 2026).

INTERNATIONAL EXAMPLE: BOGOTA'S CARE BLOCKS

Bogota's Care Blocks programme, which provides integrated care services to female caregivers across the Colombian capital, illustrates what deliberate scaling looks like in practice. What began as a concept developed by the Secretariat of Women's Affairs in 2020 grew into a network of 15 Care Blocks within two years, with plans for 20 by 2023 and 45 by 2035.

The scaling success was not accidental. Three features of the programme's design made it replicable. First, the 3 Rs framework: recognise, reduce, and redistribute the burden of care, provided a transferable conceptual model that could be adopted by new locations while accommodating local variation in services and infrastructure. Second, governance was institutionalised early through an Intersectoral Commission of Care coordinated across 13 secretariats, creating the cross-institutional structure needed to manage a city-wide rollout systematically. Third, and most significantly, the programme embedded its scaling roadmap into Bogotá's Urban Master Plan, making the programme's continuation a legal and planning requirement rather than dependent on political will or individual leadership.

This institutionalisation of the scaling logic is what transformed a promising pilot into a durable public intervention. It also illustrates the central lesson of the Scale phase: solutions that last are not those that were most innovative at the pilot stage, but those whose teams planned for sustainability from the beginning (OECD, 2023).

Source: OECD based on cited references.

**MAINSTREAMING
THE PRACTICE OF
INNOVATION:
SKILLS THAT MAKE
THE DIFFERENCE**

04

Using the tools in this guide is a starting point, not an endpoint. The most important question for any public administration is not whether a team has run one pilot, but whether innovative policymaking becomes part of how the institution works, embedded in how problems are framed, solutions are tested, and decisions are made.

Developing these behaviours builds something more than the capacity to run better pilots. OECD Research on public sector performance consistently identifies innovative capacity, the ability to experiment, learn, and adapt (Kaur, Buisman, Bekker, & McCulloch, 2022), as one of the most significant factors explaining why some public organisations deliver better outcomes than others over time (OECD, 2025). The behaviours described in this section are not only useful for the next pilot; they are the building-block competencies of an innovative mindset that improves how an organisation works across all its functions, not just those explicitly labelled as innovation.

Mainstreaming innovation is not primarily a question of method. It is a question of skills, capabilities, and behaviours. The tools in Part 3 provide structured pathways, but what people actually do with them, and what leaders do to make space for them, determines whether those pathways lead somewhere lasting. This section draws on the OECD's Forthcoming paper: Strengthening the innovative capacity of governments through skills development (OECD, forthcoming), to describe what that looks like in practice.

If you are a public servant

These four behaviours describe the day-to-day capabilities that determine whether a pilot produces genuine learning or merely goes through the motions, and whether that learning stays with the team or becomes part of how the institution works.

- **Stay open and adaptive.** Resist the pull toward the solution you already have in mind. Be willing to revise your problem definition when evidence challenges it, narrow your scope when ambition outstrips capacity, and treat uncertainty as information rather than a problem to be resolved. The Scan phase depends on this: asking better questions before proposing answers is a discipline, not just a technique.
- **Ground decisions in evidence.** Make evidence-use habitual rather than exceptional. Draw on existing studies and evaluations before designing new research. Be honest about the difference between what your evidence shows and what you are still assuming. When a team routinely reaches for the Testing Plan before building anything, evidence use stops being a feature of pilots and becomes a feature of policymaking.
- **Put people at the centre.** The people a policy is designed to affect must be involved in shaping it, not informed about it after decisions have been made. Build user engagement into work plans, protect time for it, and treat insights from interviews and observations as evidence that carries real weight. Teams that habitually involve users produce better designs, surface fewer surprises at implementation, and build stronger cases for scaling.
- **Build coalitions and share ownership.** None of the pilot cases from Bulgaria and beyond exemplified in this guide succeeded through individual effort alone.

ne. Treat stakeholder engagement as a design activity, not a communication task, building shared ownership of the problem and the evidence from the earliest possible stage. When this becomes normal practice, the question “who else needs to be involved?” is asked at the start of every initiative, not when implementation is already underway.

If you are a manager or leader

These five behaviours describe the capabilities leaders exercise day-to-day in how they run meetings, allocate time, respond to findings, and discuss challenges and objectives, which determine whether the teams around them can innovate effectively and produce improved public outcomes.

- **Protect time to innovate.** Public servants who take on innovative work alongside their routine responsibilities need a dedicated space to do it well. Running a Testing Plan, conducting user interviews, and iterating on a prototype all require focused time that cannot easily be carved out of a full workload. Protecting that time means naming it explicitly, agreeing on how many hours per week are dedicated to pilot work and holding that agreement when other priorities press. When teams have that dedicated space, the tools in this guide produce the quality of learning they are designed to generate.
- **Ask what was learned, not just what was done.** The most powerful signal a leader can send is that pilot outputs will shape real decisions. In practice, this means changing the questions asked in progress reviews: not “are you on track?” but “what did the last round of testing reveal, and how has it changed your design?” When teams see that their learning is taken seriously, they invest more in producing it.
- **Connect pilot work to organisational and government priorities.** Pilots that are linked to the institution's strategic goals from the outset are far more likely to scale. In practice, this means connecting the pilot's theory of change to existing priorities, identifying opportunities to link outputs to budget processes or policy cycles, and making the case in senior exchanges for why the work matters, not just for the pilot team but for the organisation as a whole.
- **Treat revision as a sign of learning, not a sign of failure.** When a team comes back with a redesigned prototype or a revised problem definition, the right response is “what did you learn that led to this?” rather than “why has this changed?” Leaders who consistently frame design revision as good judgment create the conditions in which teams surface problems early, exactly when they are cheapest to fix.
- **Build the conditions in which people feel safe to raise problems.** Psychological safety is the foundation on which everything else rests. It is built through small, repeated actions: crediting team members who flag risks that improve a design, acknowledging openly when an assumption turned out to be wrong, and telling the story of pilots by describing what was discovered rather than what was delivered. These behaviours signal that learning is valued, and that signal is what makes innovation possible across the whole organisation, not just within the teams already convinced.

REFERENCE

REFERENCES

- Government of Chile. (2022). **Agile Consulting LabGob: Public Innovation Methodology**. Laboratorio de Gobierno - Government Laboratory, Chile. Retrieved from <https://www.lab.gob.cl/consultoria-agil/inicio-agil>
- Kaur, M., Buisman, H., Bekker, A., & McCulloch, C. (2022). **Innovative capacity of governments: A systemic framework**. In OECD Working Papers on Public Governance. OECD Publishing, Paris. doi:10.1787/52389006-en
- OECD. (2017). **Fostering Innovation in the Public Sector**. OECD Publishing, Paris. doi:10.1787/9789264270879-en
- OECD. (2019). **Declaration on Public Sector Innovation** [OECD/LEGAL/0450]. Retrieved from OECD Legal Instruments: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0450>
- OECD. (2022). **Tackling Policy Challenges Through Public Sector Innovation: A Strategic Portfolio Approach**. In OECD Public Governance Reviews. OECD Publishing, Paris. doi:10.1787/052b06b7-en
- OECD. (2023). **Global Trends in Government Innovation 2023**. In OECD Public Governance Reviews. OECD Publishing, Paris. doi:10.1787/0655b570-en
- OECD. (2024). **Global Trends in Government Innovation 2024: Fostering Human-Centred Public Services**. In OECD Public Governance Reviews. OECD Publishing, Paris. doi:10.1787/c1bc19c3-en
- OECD. (2024). **Strengthening the Innovative Capacity of the Public Sector of Bulgaria**. In OECD Public Governance Reviews. OECD Publishing, Paris. doi:10.1787/39fa2eed-en
- OECD. (2025). **Workforce Insights from Central Governments: Findings of the 2024 OECD/EU Survey of Public Servants**. OECD Publishing, Paris. doi:10.1787/2f9080b1-en
- OECD. (2026). **Innovative Policymaking Pilots in Bulgaria - Strengthening the capacity to design and monitor public policies** - Internal Report.
- OECD. (forthcoming). **Strengthening the innovative capacity of governments through skills development**.
- Tönurist, P., & Hanson, A. (2020). **Anticipatory innovation governance: Shaping the future through proactive policy making**. In OECD Working Papers on Public Governance. OECD Publishing, Paris. doi:10.1787/cce14d80-en
- UK Government. (2025). **AI Opportunities Action Plan**. Retrieved from UK Department for Science, Innovation and Technology: https://assets.publishing.service.gov.uk/media/678639913a9388161c5d2376/ai_opportunities_action_plan_government_repsonse.pdf

ANNEX A- TOOL CANVASES

Problem Framing and Challenge Definition

Tool 1 - Scan phase

INITIAL PROBLEM STATEMENT		Complete the sentence: (User/s) need (need) due to (cause), which leads to (consequence).	
WHO IS AFFECTED AND HOW		ROOT CAUSES (WHY DOES THIS PROBLEM EXIST?)	WHAT ALREADY EXISTS (EVIDENCE AND PAST EFFORTS)
Directly affected:		Cause 1:	Evidence / data / documents:
Indirectly affected:		Cause 2:	What has been tried:
Impact (time, cost, opportunity, wellbeing):		Cause 3:	Why it fell short:
Why? (optional deeper causes)			
REFINED PROBLEM STATEMENT			
Complete and improve the sentence: (User/s) need (need) due to (cause), which leads to (consequence).			
CHALLENGE DEFINITION ("HOW MIGHT WE...")		VALIDATION AND EVIDENCE	
Complete the sentence: How might we (action) (object/system of intervention) for (user/s)?		Who tested it:	What evidence shows the problem exists and is relevant:
		Does it reflect their experience:	Key feedback:

Stakeholder mapping and feasibility assessment

Tool 2 - Scan phase

STAKEHOLDERS				
Who is affected:	Who has authority:	Who has relevant expertise:	Who might be negatively affected:	
STAKEHOLDER ASSESSMENT (FOR EACH KEY STAKEHOLDER/GROUP)				
Category	Interest	Position	Influence	Critical
Critical stakeholders (essential for success):			Key potential resistors:	
ENGAGEMENT APPROACH (FOR CRITICAL STAKEHOLDERS)		FEASIBILITY CHECK		
What do they need to support the pilot:		Senior leadership commitment:		
Likely concerns:		Staff time and capacity:		
How will you engage them:		Budget availability:		
		Regulatory environment:		
		Access to users / target groups:		
GAPS AND ADJUSTMENTS				
Key gaps identified:				
What needs to change:				
Adjustments to the pilot's challenge scope:				

Pilot concept

Tool 3 - Pilot phase

IMAGE OF THE SOLUTION		NAME, FUNCTION & VALUE		USE CASE				
Add a large, clear image or sketch that represents your solution.		Give your pilot a name and a short description.		Describe a common or meaningful situation where your solution will be used.				
KEY FEATURURES & HOW THEY WORK		PHASES OR KEY ACTIVITIES				PUBLIC & STAKEHOLDER VALUE		
Feature	Why It Matters	How It Works	Phase / Activity	Purpose	Key Actors / Tools	Stakeholder / User	Value Delivered	Why It Matters
TV PITCH		REAL INSIGHTS & EVIDENCE		EXPECTED RESULTS				
Imagine your solution is being promoted in a 30-second TV spot.		Add quotes, stories, or images that inspired or informed your idea.		Describe what changes or outcomes you expect over time.				
				Timeframe	What Will Change or Be Achieved	Indicators / Examples		
				Short-term (0- 6 months)				
				Medium-term (6-12 months)				
				Long-term (1 - 3 years)				

User and/or target group:		Pre Before the policy is experienced	During the policy intervention	Post After the policy has been delivered
USER JOURNEY - Visible to users experiencing the policy or the services it delivers				
Actions by user What the user does at each stage - applying, submitting, attending, receiving, or following up.				
Emotional / Thinking journey: What the user feels, thinks, or says at each step - their expectations, concerns, and level of trust.				
Touchpoints: People, interfaces, systems, documents, or spaces the user interacts with at each stage.				
BACKSTAGE - Not visible to users				
Internal activities / tools: Workflows, approval processes, and systems that enable each stage of the journey.				
Internal actors: Teams, units, or institutions responsible for each backstage activity, including coordination needs across organisations.				
PILOT PLANNING		HOW TO TEST THESE ASSUMPTIONS		ASSUMPTIONS ABOUT USERS
The method to test each assumption - interviews, surveys, observation, usability testing, role-play, or prototype		What the team is assuming about user behaviour, motivations, or access at each stage.		

Testing Plan

Tool 5 - Pilot phase

ASSUMPTION TO TEST	ACTIVITY AND PROTOTYPE
<p>Main assumptions to start our testing. Example: Simulation of service delivery at the citizen service counter to observe user interaction with forms, waiting times, and guidance received.</p>	<p>Prototype and activities you will carry to test the assumption Example: Simulation of service delivery at the citizen service counter to observe user interaction with forms, waiting times, and guidance received.</p>
OBJECTIVE	
<p>What are you expecting to gain, validate or learn? Example: Understand how people interact with the prototype and whether it is clear and easy to use.</p>	
RESOURCES	DEADLINES / RESPONSIBLE PERSONS
<p>What human, tech, policy or monetary resources are needed for the test Example: Who will lead the testing, how many hours it will require, materials or tech we need for the test.</p>	<p>Who will be responsible of what and when. Example: Define responsible persons, deadlines, and next steps.</p>

Low-fidelity prototyping canvas

Tool 6 - Pilot phase

WHAT TO BUILD	WHAT TO TEST	WHO TO TEST WITH	WHAT YOU LEARNED
<p>Format: The type of prototype you will build — sketch, flowchart, scenario, role-play script, or slide deck.</p> <p>Fidelity: Keep it as rough as possible. The rougher it is, the more honest the feedback.</p>	<p>Assumption: The specific assumption from your Testing Plan this prototype is designed to test.</p> <p>Learning objective: What a successful test would tell you — in one sentence.</p>	<p>Participants: Five to eight real users from your target group. Not colleagues already familiar with the concept.</p> <p>Roles: Who facilitates, who observes, who takes notes.</p>	<p>What worked: What users responded to positively or navigated without difficulty.</p> <p>What was confusing: Where users got stuck, asked questions, or made unexpected choices.</p> <p>What to change: The one adjustment to make before testing again.</p> <p>Next step: Revise and retest, or move forward.</p>

Theory of change

Tool 8 - Pilot phase

PROJECT PURPOSE		MONITORING INDICATORS	
PROBLEMS / NEEDS			
INPUTS / ACTIVITIES			
PRODUCTS / SERVICES		OUTCOME INDICATORS	
INTERMEDIATE RESULTS			
FINAL RESULTS			
ASSUMPTIONS THAT UNDERPIN THE ABOVE			

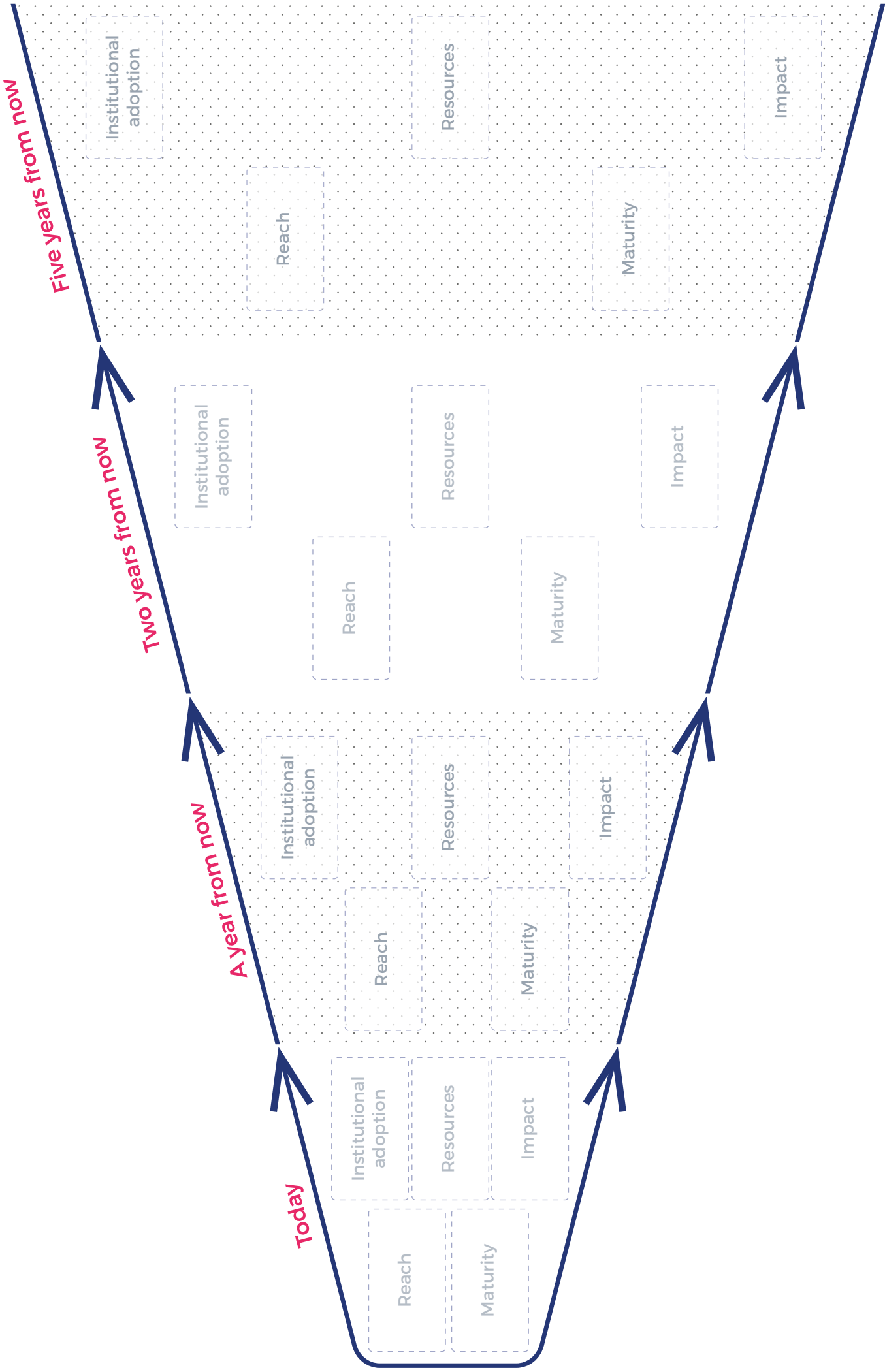
Pilot evaluation framework

Tool 9 - Scale phase

EVALUATION QUESTIONS		INDICATORS	BASELINES
What are the two or three most important things you want to know at the end of the pilot? Link each question directly to the assumptions in your theory of change.		<p>Monitoring indicators: What will you count or measure to track whether implementation is proceeding as planned? Link each indicator to a specific activity or output.</p> <p>Outcome indicators: What will you measure to assess whether the pilot is creating the intended change? Link each indicator to a specific intermediate or final outcome from your theory of change.</p>	What is the current state of each indicator before the pilot begins? Where will you find this data, administrative records, previous surveys, or user research from the Scan phase?
DATA COLLECTION		DECISION THRESHOLDS	
<p>Method: How will you collect data for each indicator, administrative records, surveys, interviews, observation, or service data?</p> <p>Frequency: When will you collect data, before the pilot begins, at a mid-point review, and at the end?</p>		<p>Responsible person: Who is accountable for collecting and recording the data for each indicator?</p>	
		What level of evidence would be sufficient to recommend scaling? Define this before you see the results.	

Scaling funnel

Tool 10 - Scale phase



INNOVATING IN
BULGARIA'S PUBLIC
ADMINISTRATION:
**A PRACTICAL GUIDE
TO DESIGNING AND
RUNNING POLICY PILOTS**



Funded by
the European Union