

UX evaluation templates

Three structured instruments for evaluating public services: auditing what exists, deciding what to fix, and scoring what you build.

What is in this pack

Three templates that follow the lifecycle of any service improvement: diagnose the current state, decide what is worth solving, then test the new design against the same vocabulary you started with.

1

Current Service Evaluation

Audit an existing service against the three pillars (effectiveness, efficiency, satisfaction), the ten principles of usable services, and the citizen journey from discovery to post-decision. Output: a defensible diagnostic report.

2

Problem Priority Assessment

Score one candidate problem on five importance dimensions (reach, severity, equity, hidden agency cost, trust impact) and four tractability dimensions (complexity, time, dependencies, capability). Output: a decision — do now, invest, defer, or drop.

3

Prototype UX Score

Score a tested prototype across the same three pillars, with per-task metrics, a severity-rated issue log, and a coverage check for hard-to-reach and assisted-digital cases. Output: ready, iterate, rethink, or kill.

How they fit together

Use Template 1 to understand where a service stands today. Run Template 2 on each candidate problem the evaluation surfaces, to decide what to tackle first. After redesign and testing, use Template 3 to verify the new design works — ideally against the same three-pillar scoring you started with, so the before-and-after comparison is direct.

All three templates share the vocabulary from the Service Usability workshop. They work as standalone instruments without the workshop, but the depth of judgment they support is higher for teams who have done the workshop together.

TEMPLATE 1

Current Service Evaluation

For auditing an existing public service against the three pillars and ten principles.

How to use this. Allow 60 to 90 minutes. Best filled in by two or more evaluators, ideally after a Service Walk on the actual live service. Disagreement between evaluators is signal, not noise — record both views.

SERVICE PROFILE

Service name _____

Owning team _____

Primary purpose _____

Channels evaluated _____

Evaluator(s) _____

Date _____

THE THREE PILLARS AT A GLANCE

PILLAR	SCORE	EVIDENCE AND BIGGEST GAP
EFFECTIVENESS Can the citizen complete the task and is the outcome correct?	Score (circle one) 1 2 3 4 5	_____ _____
EFFICIENCY How much time, effort, and re-work did it cost?	Score (circle one) 1 2 3 4 5	_____ _____
SATISFACTION Did the citizen feel respected? Would they return without dread?	Score (circle one) 1 2 3 4 5	_____ _____

TEN PRINCIPLES AUDIT - ONE ROW PER PRINCIPLE

PRINCIPLE	STATUS	SEVERITY	WHERE / EVIDENCE
One thing per page	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>
Use citizen language	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>
Show eligibility before personal details	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>
Save and resume by default	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>
Make status visible after submission	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>
Design for the assisted-digital case	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>
Explain what you will do with their information	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>
Confirm submission in plain English	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>
Design the error message before the validation	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>
Test the unhappy path	<input type="checkbox"/> OK <input type="checkbox"/> Partial <input type="checkbox"/> Violated <input type="checkbox"/> N/A	<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical	<hr/> <hr/>

JOURNEY-STAGE FINDINGS

STAGE	WHAT WORKS / WHAT DOESN'T / BIGGEST ISSUE
Discovery How do citizens find out the service exists?	<hr/> <hr/>
Eligibility Do they understand if it applies to them? Do some self-exclude wrongly?	<hr/> <hr/>
Application The form, the questions, the proof, the upload.	<hr/> <hr/>
Decision How the outcome is delivered and explained.	<hr/> <hr/>
Post-decision Status, follow-up, appeals, recourse.	<hr/> <hr/>

TOP 5 ISSUES

Promote the issues most worth solving to Template 2 for priority scoring.

#	ISSUE	WHO IS AFFECTED	SEVERITY
1	<hr/>		<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical
2	<hr/>		<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical
3	<hr/>		<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical
4	<hr/>		<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical
5	<hr/>		<input type="checkbox"/> Low <input type="checkbox"/> Med <input type="checkbox"/> High <input type="checkbox"/> Critical

TEMPLATE 2

Problem Priority Assessment

For deciding whether and how urgently a problem is worth solving.

How to use this. Score one problem at a time. Run multiple sheets for multiple problems, then compare totals. The five Importance criteria and four Tractability criteria are deliberately tuned for public services: standard product frameworks (RICE, ICE) miss equity, trust impact, and the hidden cost of inaction — which often dominate the real picture.

PROBLEM STATEMENT

One paragraph. What is the problem, whom does it affect, and what happens to them?

IMPORTANCE · IS THIS WORTH SOLVING?

1. REACH how many citizens per year?

1 = under 1K · 2 = 1K–10K · 3 = 10K–100K · 4 = 100K–1M · 5 = over 1M

Score (1–5): _____ EVIDENCE _____

2. SEVERITY FOR THE CITIZEN what is the worst outcome?

1 = mild inconvenience · 2 = time cost / friction · 3 = repeated effort · 4 = blocked from entitlement · 5 = real harm

Score (1–5): _____ EVIDENCE _____

3. EQUITY is it concentrated on vulnerable groups?

1 = random · 2 = slight skew · 3 = meaningfully concentrated · 4 = severely concentrated · 5 = creates exclusion

Score (1–5): _____ EVIDENCE _____

4. HIDDEN AGENCY COST call volume, appeals, manual work?

1 = negligible · 2 = moderate · 3 = significant · 4 = budget driver · 5 = wouldn't exist if fixed

Score (1–5): _____ EVIDENCE _____

5. TRUST IMPACT does this erode trust in the agency or state?

1 = invisible · 2 = occasional complaint · 3 = recurring criticism · 4 = known reputational issue · 5 = cited publicly as failure

Score (1–5): _____ EVIDENCE _____

IMPORTANCE TOTAL (sum of 5 scores, max 25) _____

TRACTABILITY · CAN WE ACTUALLY SOLVE IT?

1. COMPLEXITY OF THE FIX what kind of change is needed?

5 = UI / content · 4 = team process · 3 = cross-team process · 2 = IT system · 1 = policy or legal

Score (1–5): _____ EVIDENCE _____

2. TIME TO SHIP A MEANINGFUL IMPROVEMENT first measurable improvement in...

5 = < 4 weeks · 4 = 1–3 months · 3 = 3–6 months · 2 = 6–12 months · 1 = > 12 months

Score (1–5): _____ EVIDENCE _____

3. DEPENDENCIES who else has to act?

5 = none · 4 = one other team · 3 = cross-functional inside agency · 2 = cross-agency · 1 = external (legislature, vendor)

Score (1–5): _____ EVIDENCE _____

4. CAPABILITY AVAILABLE do we have the people?

5 = capability has headroom · 4 = capability is stretched · 3 = partial · 2 = gap, needs hiring · 1 = does not exist in org

Score (1–5): _____ EVIDENCE _____

TRACTABILITY TOTAL (sum of 4 scores, max 20) _____

DECISION

High importance (≥17) + high tractability (≥14)	DO NOW. Top-of-backlog for this quarter.
High importance (≥17) + low tractability (≤8)	INVEST. Build capability, unblock dependencies, partner up. Worth a 6–12 month horizon.
Low importance (≤10) + high tractability (≥14)	DEFER. Easy is not the same as worth doing. Don't burn cycles on trivial wins.
Low importance (≤10) + low tractability (≤8)	DROP. Document the decision and move on.
Anywhere in the middle	Judgement call. Use justification box below to record the call and its reasoning.

RECOMMENDATION AND JUSTIFICATION

Recommended by _____ Date _____

TEMPLATE 3

Prototype UX Score

For scoring a tested prototype before deciding whether to ship, iterate, or rethink it.

How to use this. Fill in after testing the prototype with at least 5 representative users — ideally including someone from a hard-to-reach group. Fewer than 5 users, or colleagues only, makes this a directional rather than validated evaluation; mark it as such.

PROTOTYPE PROFILE

Prototype / feature _____

Version / date _____

HYPOTHESIS BEING TESTED

"I am testing whether X will help specified user do Y."

WHAT THE PROTOTYPE TESTED

WHAT IT DELIBERATELY DID NOT TEST

PARTICIPANTS

#	SEGMENT	DONE?	TIME	SEQ (1–7)	KEY OBSERVATION
P1		<input type="checkbox"/> Y <input type="checkbox"/> N <input type="checkbox"/> Partial			
P2		<input type="checkbox"/> Y <input type="checkbox"/> N <input type="checkbox"/> Partial			
P3		<input type="checkbox"/> Y <input type="checkbox"/> N <input type="checkbox"/> Partial			
P4		<input type="checkbox"/> Y <input type="checkbox"/> N <input type="checkbox"/> Partial			
P5		<input type="checkbox"/> Y <input type="checkbox"/> N <input type="checkbox"/> Partial			

THREE-PILLAR SCORE FOR THE PROTOTYPE

Score the prototype 1–5 on each pillar, evidenced by observed behaviour during the sessions.

PILLAR	SCORE	EVIDENCE
EFFECTIVENESS observed completion + correctness of outcome	1 2 3 4 5	_____ _____ _____
EFFICIENCY observed time + observed errors	1 2 3 4 5	_____ _____ _____
SATISFACTION participants' SEQ + post-task comments	1 2 3 4 5	_____ _____ _____

ISSUE LOG

Severity: **Cosmetic** = typo, label tweak · **Minor** = friction, still completes · **Major** = some users fail or get wrong outcome · **Critical** = systematic failure.

#	ISSUE	SEVERITY	# OF Ps	RECOMMENDED RESPONSE
1		<input type="checkbox"/> Cos <input type="checkbox"/> Min <input type="checkbox"/> Maj <input type="checkbox"/> Crit		
2		<input type="checkbox"/> Cos <input type="checkbox"/> Min <input type="checkbox"/> Maj <input type="checkbox"/> Crit		
3		<input type="checkbox"/> Cos <input type="checkbox"/> Min <input type="checkbox"/> Maj <input type="checkbox"/> Crit		
4		<input type="checkbox"/> Cos <input type="checkbox"/> Min <input type="checkbox"/> Maj <input type="checkbox"/> Crit		
5		<input type="checkbox"/> Cos <input type="checkbox"/> Min <input type="checkbox"/> Maj <input type="checkbox"/> Crit		
6		<input type="checkbox"/> Cos <input type="checkbox"/> Min <input type="checkbox"/> Maj <input type="checkbox"/> Crit		

COVERAGE CHECK

If you cannot tick a box, mark this evaluation as *partial* in the verdict below.

- | | | |
|------------------------------|-----------------------------|---|
| <input type="checkbox"/> Yes | <input type="checkbox"/> No | The prototype worked for the bullseye user. |
| <input type="checkbox"/> Yes | <input type="checkbox"/> No | It worked for at least one ring-1 audience. |
| <input type="checkbox"/> Yes | <input type="checkbox"/> No | The unhappy path was tested (eligibility failure, validation, time-out, user error). |
| <input type="checkbox"/> Yes | <input type="checkbox"/> No | The assisted-digital case was tested (one user helping another through the screen). |
| <input type="checkbox"/> Yes | <input type="checkbox"/> No | Reading age was verified at grade 9 or lower. |
| <input type="checkbox"/> Yes | <input type="checkbox"/> No | At least one participant came from a hard-to-reach group (older, low digital literacy, non-native speaker, claimant). |

VERDICT

Pick one. Make the team's reasoning explicit; this is the artefact that will be referenced when the next prototype is briefed.

- | | |
|--------------------------|---|
| <input type="checkbox"/> | READY. Move to next development phase as-is. |
| <input type="checkbox"/> | ITERATE. Specific changes needed before next phase (list below). |
| <input type="checkbox"/> | RETHINK. The approach is wrong; redirect (direction below). |
| <input type="checkbox"/> | KILL. Don't pursue further; record reasoning so the learning isn't lost. |

REASONING AND NEXT ITERATION BRIEF

Signed off by _____ **Date** _____