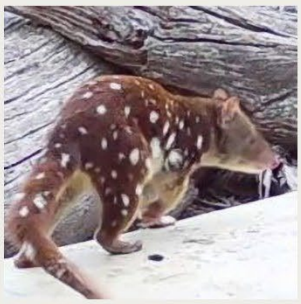
Healthy Forests
Foundation.org

Conservation Plan Scorecard

Assessment method






How we assess plans for threatened species and ecological communities against 15 standards, and how anyone can apply the scorecard themselves.

Quollity SCORECARD

October 2026

Conservation Plan Scorecard: assessment method

This document sets out how the Healthy Forests Foundation assesses conservation plans with the Conservation Plan Scorecard: the 15 standards, what earns each rating, how ratings become a grade, and how assessments are checked, published and corrected. It is published so that anyone can apply the scorecard themselves and test our assessments against it.

Contents

Conservation Plan Scorecard: assessment method....1
Purpose and scope....3
Basis of the standards....3
Rating scale and grades....5
Rating criteria....7
What is to be conserved, and to what level?....7
Where does effort matter most?....8
Which actions come first, and at what cost?....9
How will success be judged, and who is responsible?....9
Counting rules and definitions....11
Assessment procedure....12
Limitations and disclosure....12
Version history....13
Sources....13
Appendix: Suggested AI prompt....14

Purpose and scope

The scorecard tests one thing: whether a conservation plan, as published, can guide decisions. It asks whether the plan answers four questions:

1. What is to be conserved, and to what level?
2. Where does effort matter most?
3. Which actions come first, and at what cost?
4. How will success be judged, and who is responsible?

What it assesses. The scorecard is designed for plans for a single threatened species or ecological community, such as national recovery plans, state recovery plans and conservation advices. It assesses the plan document as published, at the version identified in the assessment. A draft released for comment is assessed as a draft and labelled as one.

What it does not assess. The scorecard does not assess:

- whether the plan is being implemented, or whether the species is recovering
- the scientific merit of individual studies the plan relies on, beyond whether its key claims are traceable to them
- whether the species should be listed, or at what category
- the performance of the agency or people who wrote the plan.

Threat abatement plans and multi-species plans can be assessed, but several standards (notably irreplaceability and adequacy) were calibrated on single-species plans. Assessments of those plan types are labelled as provisional until the method has been tested on them.

Basis of the standards

Ten of the 15 standards come from the core concepts of systematic conservation planning reviewed by Kukkala and Moilanen (2013); 3 reflect elements of the planning process that the review discusses; 2 are the Foundation's own additions.

The review covers twelve concepts: adequacy, complementarity, comprehensiveness, effectiveness, efficiency, flexibility, irreplaceability, replacement cost, representation, representativeness, threat and vulnerability. Two are omitted. Representativeness and replacement cost describe properties of a network of protected areas, and do not translate to a plan for one species or community.

The descriptions below are the Foundation's paraphrases and adaptations. The review deals with spatial prioritisation across many features; the scorecard applies its concepts to the narrower question of whether a single plan can direct effort.

Standard	Origin	How the scorecard applies it
Comprehensiveness	Review concept	The plan covers every unit and feature that needs conserving.
Adequacy	Review concept	The plan sets a level of conservation high enough to secure long-term survival.
Representation and persistence	Review concept (representation); the persistence test is the Foundation's emphasis	The plan separates where the species occurs from where it is likely to persist.
Threat	Review concept	Threats are rated on evidence in a way that ranks them.
Vulnerability	Review concept	The plan shows which places are most likely to be lost.
Irreplaceability	Review concept	The plan identifies places that cannot be substituted.
Complementarity	Review concept, adapted	The plan considers how its actions interact with each other and with other species. In the review, complementarity concerns how sites add to one another; the scorecard applies the same idea to actions.
Efficiency	Review concept	Priorities reflect the benefit gained for the cost. The review notes that efficiency in spatial prioritisation frequently means cost efficiency, the ratio of benefits to costs.
Flexibility	Review concept	The plan allows more than one way of meeting its targets.
Effectiveness	Review concept	Success can be measured against targets and used to adapt management.
Clear, operable definitions	Planning process	The review is concerned with linguistic uncertainty in planning; the scorecard tests whether the plan's decision-triggering terms can be applied consistently.
Engagement	Planning process	Stakeholders, including First Nations, shape and deliver the plan.
Social and economic considerations	Planning process	The plan assesses impacts on communities and land uses.
Evidence quality	Foundation addition	Key claims and ratings are traceable to published sources.
Accountability	Foundation addition	Each action has a lead party, timeframe and funding.

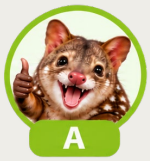
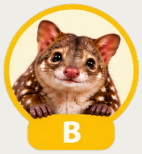
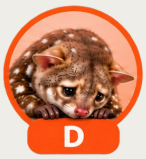
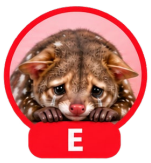
Comprehensiveness, adequacy and representativeness also underpin Australia's forest reserve criteria under the National Forest Policy Statement (1992) and the JANIS criteria (1997).

Rating scale and grades

Each standard earns 0, 1 or 2 points; the total out of 30 sets a grade, and the five decision-critical standards can cap that grade.

Rating	Points	Meaning
Met	2	The plan satisfies the standard in a form that can guide a decision.
Partly met	1	The plan addresses the standard, but not in a form that can be applied or assessed.
Not met	0	The plan does not address the standard.

The percentage is the total divided by 30, not rounded. A plan scoring 23 (76.7%) is a B+, not an A.

Grade	Score	Meaning
	90% or more (27 to 30)	Exemplary. A model for other plans.
	80 to 89% (24 to 26)	Strong. A clear guide to decisions, with minor gaps.
	70 to 79% (21 to 23)	Sound. Guides decisions, though some standards are only partly met.
	60 to 69% (18 to 20)	Adequate. Usable, but with significant gaps.
	50 to 59% (15 to 17)	Weak. Describes the species and threats but gives limited guidance.
	35 to 49% (11 to 14)	Poor. Cannot reliably guide decisions.
	Below 35% (0 to 10)	Not fit for purpose as a decision document.

Decision-critical caps. Adequacy, irreplaceability, efficiency, effectiveness and accountability answer the four questions directly. A plan cannot earn a good grade on description alone while failing them:

- one decision-critical standard not met: grade no higher than B
- 2 not met: no higher than C
- 3 or more not met: no higher than D.

Caps apply only to standards rated Not met. A Partly met rating does not trigger a cap. Where a cap lowers a grade, the assessment says so (for example, "B, capped from B+").

Rating criteria

Each standard has a written test for Met, Partly met and Not met, so that 2 assessors reading the same plan reach the same rating. Decision-critical standards are marked *. Where a plan falls between 2 levels, the lower rating applies and the assessment explains why. A plan falls between 2 levels when it meets part of the higher level's test but not all of it, or when its wording fits both levels. Do not lower a rating for anything the test does not ask for. Evidence can come from anywhere in the plan, not only the section where you would expect to find it.

What is to be conserved, and to what level?

Standard	Met (2)	Partly met (1)	Not met (0)
Comprehensiveness	Names or maps the units to be conserved (populations, subpopulations, evolutionarily significant units or management units) and the habitat and life-stage needs of each, across the whole range.	Describes the species and its range but treats it as a single unit without justification, or defers the definition of units to future research.	Defines nothing to be conserved beyond the species name.
Adequacy*	Sets a quantified target linked to long-term persistence, such as a rate of increase, a minimum population size, a number of populations of stated viable size, or an amount of habitat.	Sets a direction of change tied to something specific: named units, a habitat retention threshold, or a genetic exchange rule.	Sets only a direction of change ("stable or increasing", "a measurable increase", "a positive trend").
Representation and persistence	Distinguishes where the species has been recorded from where it is likely to persist, at named places (for example, identifying populations unlikely to persist, or core areas reliably occupied across conditions).	Records distribution and populations, or states criteria for persistence, without applying them to places.	Gives distribution only, with no consideration of persistence.

Where does effort matter most?

Standard	Met (2)	Partly met (1)	Not met (0)
Threat	Rates each threat (for example, by likelihood and consequence) on cited evidence, with no more than half of the threats in the highest category.	Rates threats but places more than half in the highest category, or describes threats on evidence without rating them.	Lists threats without rating or evidence.
Vulnerability	Identifies the named places or populations most likely to be lost, and why.	Describes exposure to threats across the range, or as a proportion of habitat, without locating it.	Does not consider which places are most likely to be lost.
Irreplaceability*	Identifies the places that cannot be substituted, defined so they can be located (mapped, declared or bounded), and treats their loss as unacceptable or uncompensable.	States criteria for important populations or habitat, or names places without boundaries, or distinguishes some units as more important than others.	Defines critical habitat so broadly that any area with one listed attribute qualifies; declares all occupied areas equally important with nothing further; or states that critical habitat cannot be delineated or is left to future research.
Complementarity	Considers how actions interact with each other and with other species, including conflicts and unintended consequences, not only shared benefits.	Lists species or communities that will benefit, or notes a single interaction.	Does not consider other species or interactions between actions.

Which actions come first, and at what cost?

Standard	Met (2)	Partly met (1)	Not met (0)
Efficiency*	Costs each action and sets priorities by expected benefit relative to cost, through a stated method that weighs one against the other.	Costs each action but sets priorities on urgency alone, or does not prioritise discriminatingly; or costs only some actions.	Does not cost actions, however they are ranked. A commitment to analyse cost-effectiveness later does not count until the analysis is in the plan.
Flexibility	States alternative ways or places of meeting targets, with criteria for choosing between them. The choice itself can be left until later.	Refers to adaptive or site-tailored management without stating alternatives.	Prescribes a single approach, or gives no basis on which alternatives could be judged.
Social and economic considerations	Assesses the likely costs and benefits of the plan for communities, industries and land uses. A description in words is enough.	Names affected industries or land uses without assessing impacts.	Has no consideration, or only a list of stakeholders.

How will success be judged, and who is responsible?

Standard	Met (2)	Partly met (1)	Not met (0)
Effectiveness*	Sets measurable performance criteria against a stated baseline, specifies or has in place monitoring able to detect change, and states how results will be reviewed and used to adapt management.	Proposes monitoring, or sets criteria without a stated baseline.	Has no measurable criteria.

Standard	Met (2)	Partly met (1)	Not met (0)
Clear, operable definitions	Defines the terms that trigger decisions (critical habitat, important population, thresholds) in measurable terms, or in a glossary sufficient for consistent application.	Quantifies some decision-triggering terms but leaves others vague (for example, "high densities" or "medium or large sites").	Leaves decision-triggering terms undefined.
Engagement	Documents how stakeholders, including First Nations, were involved in developing the plan, and gives them defined roles in delivering it.	Documents one but not the other, or documents consultation without First Nations involvement.	Documents neither.
Evidence quality	Key claims (population size and trend, distribution, habitat requirements) and threat ratings are traceable to published sources.	Key claims or ratings rest mainly on personal communications or unsourced expert opinion.	Key claims are unsupported.
Accountability*	Each action has a single lead party, a timeframe, and a funding source or indicative cost.	Responsible parties are named for actions (a list of potential partners is enough), but the lead, timeframe or funding is missing.	No responsible party is named for actions, whether or not timeframes are given.

The grouping by question is for readability; it does not change how any standard is scored. Some standards sit where they do for a reason that is not obvious. Flexibility and social and economic considerations sit with cost because they bear on how, and at whose expense, effort is spent. Definitions, engagement and evidence quality sit with judging success because they determine whether a plan's claims can be tested and who can test them.

Counting rules and definitions

Several ratings depend on counts or terms that need a fixed meaning. These apply to every assessment.

- **Action.** The lowest-level numbered action in the plan's action tables. Sub-points within an action are not counted separately. Where a plan groups actions under strategies without numbering them, each separately described action is counted.
- **Threat.** Each distinct threat named in the plan's threat rating table or matrix. A threat appearing in 2 cells of a matrix is counted once, and the inconsistency is noted.
- **Highest category.** The top category of the plan's own priority or risk scale (for example, Priority 1, Very High or High where High is the top level).
- **More than half.** Strictly more than 50% of actions or threats. A plan with exactly half in the top category passes.
- **Separate populations or assessments.** Where a plan rates threats or prioritises actions separately for different populations, each set is tested, and the standard is met only if every set passes.
- **Costed.** An indicative dollar cost is given for each action, or for each group of actions where the group is the unit of delivery. Costs stated only as "core business" or "existing programs" do not count as costs. An action the plan says is funded under another costed action counts as costed.
- **Cost-effectiveness method.** Any stated approach in which priorities are set by comparing expected benefit with cost, such as ranking actions by benefit per dollar, or a structured decision process whose results are reported in the plan.
- **Key claims.** Population size and trend, distribution and area of occupancy, habitat requirements, and the evidence behind threat ratings.
- **Mainly personal communications.** Where personal communications or unsourced expert opinion are the only support for one or more key claims, or account for a large share of the citations in the sections setting out key claims. The assessment records the count. Count them across the whole plan, excluding the bibliography.
- **Lead party.** A single named organisation responsible for delivery. A list of several organisations counts as a lead only if one is identified as the lead.
- **Timeframe.** A date, financial year or duration specific to the action. "Ongoing" is not a timeframe.
- **Baseline.** A stated value, or a stated date and method from which a value has been or will be measured, against which a performance criterion is judged. A reference to a baseline that the plan does not give, and does not say how to obtain, does not count.

Spatial standards (representation and persistence, vulnerability, irreplaceability) are judged against the species' range. A species confined to one locality can meet them by mapping that locality; a wide-ranging species must show where, within its range, effort matters most.

Assessment procedure

Every published assessment follows the same 6 steps, and every rating is tied to the section of the plan it relies on. An assessment can be carried out by a person or with the help of an AI tool. Where an AI tool is used, a person checks the result before it is used or published (step 6). A suggested prompt for AI tools is in the Appendix.

1. **Identify the plan.** Record the title, status (draft or final), date, version and the URL from which it was obtained. Keep a copy of the version assessed. Record who carried out the assessment and whether an AI tool was used. If one was, record which tool and the date.
2. **Read the whole plan.** The assessor reads the plan in full, including tables, appendices and any documents the plan states are part of it. Documents the plan merely cites are not assessed. Where an AI tool is used, confirm it has read the whole plan. Some tools cut off long documents without saying so.
3. **Record evidence for each standard.** For each of the 15 standards, the assessor records the rating, the section, table or page relied on, a short paraphrase or quotation of the relevant text, and the rule applied. Where a rating rests on absence (for example, no costs), the assessor records where they looked. For a Word or web file without fixed page numbers, give the section, table or action number instead of a page.
4. **Apply the counting rules.** Counts of actions, threats and priority levels are made directly from the plan's own tables, and recorded. Where an AI tool makes the counts, a person checks them against the tables.
5. **Calculate the grade.** Total the points, calculate the percentage, apply the decision-critical caps, and record any cap applied.
6. **Check the assessment.** Where an AI tool was used, a person checks the assessment against the plan before it is used or published. At a minimum, check every quote and page reference, every count, every rating that rests on the absence of evidence, and any decision-critical standard rated Not met. The person who checks the assessment is responsible for it.

New versions of a plan. When a plan is replaced or a draft is finalised, the new version is assessed as a new assessment. Earlier assessments remain published, labelled with the version they assessed.

Limitations and disclosure

The scorecard measures whether a plan can guide decisions, not whether a species is recovering, and it has known biases that readers should weigh.

- **Document, not delivery.** A well-written plan may be poorly implemented, and a weak plan may sit alongside good work on the ground. A grade says nothing about either.
- **Narrow-range species have an advantage.** Spatial standards are easier to meet for a species confined to one place than for one spread across a continent. The method accepts

this, because a plan for a wide-ranging species still needs to say where effort matters most, but grades should be compared with this in mind.

- **Efficiency is a demanding test.** In testing, no plan met the efficiency standard, because none set priorities by benefit relative to cost. The standard still separates costed from uncosted plans, but it currently caps or lowers every grade it touches.
- **Judgement remains.** Written criteria narrow, but do not remove, the judgement involved in rating. The evidence record, second assessment and right of reply exist so that judgements can be seen and challenged.
- **AI tools need checking.** AI tools can work through a large amount of information quickly, which helps us decide where to focus our effort across the many plans governments produce. They can also misread a plan, miss evidence or state things with more confidence than the evidence supports. That is why step 6 applies to every AI-assisted assessment.
- **Calibration.** The method was calibrated on twelve national recovery plans for single species. It has not yet been tested on threat abatement plans, ecological community plans or state plans.
- **Weighting.** All 15 standards carry equal points. The decision-critical caps are the only weighting. Other weightings are defensible; we chose the simplest that stops a plan grading well on description alone.

Disclosure. The Healthy Forests Foundation is a not-for-profit focused on native forest restoration, carbon policy and biodiversity markets. It operates a fee-for-service model alongside its charitable activities. Each published assessment states whether an AI tool was used.

Version history

Version 1.0 is the first published method; it incorporates changes made after testing on 12 national recovery plans.

Sources

- Kukkala AS and Moilanen A (2013) 'Core concepts of spatial prioritisation in systematic conservation planning', *Biological Reviews*, 88:443–464, doi:10.1111/brv.12008.
- Commonwealth of Australia (1992) *National Forest Policy Statement*.
- JANIS (1997) *Nationally Agreed Criteria for the Establishment of a Comprehensive, Adequate and Representative Reserve System for Forests in Australia*.

The scorecard, its additional standards, rating criteria, counting rules and grading system are the Foundation's own. Descriptions of concepts drawn from Kukkala and Moilanen are the Foundation's paraphrases.

Appendix: Suggested AI prompt

Anyone can use an AI tool to help apply the scorecard. This is the prompt we suggest. It is written so that different AI tools reach the same ratings on the same plan.

To use it, give the AI tool 2 documents: this method and the plan you want to assess. Copy the prompt below into the tool and replace [plan name] with the name of the plan. Whatever the tool produces is a draft. A person must check it against the plan before it is used or shared (step 6 of the Assessment procedure).

I need your help to assess a conservation plan using the Conservation Plan Scorecard, published by the Healthy Forests Foundation.

I have given you 2 documents:

1. The scorecard method: "Conservation Plan Scorecard: assessment method"
2. The plan to assess: [plan name]

RULES

- Use only these 2 documents. Do not use outside knowledge about the species, the plan or the scorecard. Do not assess documents the plan only cites.
- Follow the method exactly as written: the Assessment procedure, rating criteria, counting rules and definitions, and decision-critical caps.
- Apply the method's rule for plans that fall between 2 levels (see the paragraph above the rating criteria tables). Do not lower a rating for anything the test does not ask for.
- Search the whole plan for every standard, not only the section where you would expect the evidence.
- Back every rating with evidence from the plan and a short quote. Never invent a quote, page number or figure. If you cannot find evidence, say so and list where you looked.
- Give references as the method's Assessment procedure (step 3) describes. For a Word or web file, use the section heading, table or action number, not a page number.
- If you are unsure, say so. Do not guess.

STEP 1: IDENTIFY THE PLAN

Record the plan's title, status (draft or final), date and version, and the web address if it appears in the document. Record the AI tool used and today's date.

STEP 2: CONFIRM YOU HAVE READ THE WHOLE PLAN

List the plan's main sections, tables, figures and appendices. If either document is cut off, too long or unreadable, stop and tell me which parts you cannot see. Otherwise, continue. List every action number in the recovery action tables, from first to last, and give the total.

STEP 3: RATE EACH STANDARD

Work through all 15 standards in the order they appear in the method's rating criteria tables. For each standard:

1. List the separate parts of the Met test.
2. For each part, search the whole plan and mark it present or absent, with a quote and location.
3. Rate Met only if every part is present. If not, test Partly met the same way, then Not met.
4. Check in both directions. For a rating below Met, search again for evidence of the higher rating. For a Met rating, search for anything that rules it out.

Then present your results as a table with these columns:

- Standard
- Rating (Met, Partly met or Not met) and points (2, 1 or 0)
- Where in the plan
- Evidence (short quote or close paraphrase)
- Parts of the test met and missed
- Method wording that decided the rating (quote it)
- Confidence (high, medium or low)

Where a rating rests on the absence of evidence, say where you looked.

STEP 4: APPLY THE COUNTING RULES

Show every count you made, which table or section you counted from, and how you counted. Include at least:

- the number of actions
- the number of actions at each priority level, and whether more than half are at the highest level
- the number of actions costed, including any that count because the plan says they are funded under another costed action
- the number of threats in the plan's threat rating table, and how many are in the highest category (or say there is no such table)
- the number of actions with a single lead party
- the number of personal communications and similar citations across the whole plan, excluding the bibliography, and which key claims rest only on them.

STEP 5: CALCULATE THE GRADE

Show:

- total points out of 30
- percentage (total divided by 30, not rounded)
- grade before caps, using the method's grade table
- any decision-critical standards rated Not met, and the cap that applies
- final grade.

STEP 6: PREPARE FOR THE HUMAN CHECK

The method requires a person to check this assessment before it is used or published. To help them, list:

- every quote and reference you used
- every count you made
- every rating that rests on the absence of evidence
- every decision-critical standard rated Not met
- any rating you gave low confidence, or that was close to the line between 2 levels, and what the final grade would be if that rating moved one level.

Then give the plan's main strengths and main gaps (up to 3 each), each linked to a standard.

If you run out of space, stop at the end of a step or standard and tell me where you stopped. I will ask you to continue.

End with this sentence: "This is a draft assessment produced by an AI tool. Under step 6 of the method, a person must check it against the plan before it is used or shared."