



SERIES THE RECORD · THREE PAPERS · OPEN LICENCE

The Inherited *Default.*

Working notes on deciding when several teams independently arrived at the same practice

PAPERS

- | | | |
|-----------|---|----------------------|
| PAPER 001 | Rarity is a property of the corpus | METHOD · 8 MIN |
| PAPER 002 | A convention is a count, not a judgement | CONSTRUCTION · 8 MIN |
| PAPER 003 | What a project that does not exist yet may be told | LIMITS · 8 MIN |

ISSUED

three papers · published 18 September 2026 · CC BY 4.0
digital1.foundation/articles/the-inherited-default/

Stichting Digital One Foundation i.o. · Amsterdam

These notes describe work contributed to an entity in formation.

Contents

001

Rarity is a property of the *corpus*

METHOD • 8 MIN • PUBLISHED 18 SEPTEMBER 2026

Twenty planted conventions came back as two groups of up to 41 members, with no error raised. Recovery tracks one weight ratio, and nothing else.

§1The failure arrived as a well-formed answer3

§2A planted corpus, because the real one is private4

§3Recovery tracks one ratio and nothing else5

§4A group judged against itself6

§5What the planted corpus cannot settle7

References8

002

A convention is a count, not a *judgement*

CONSTRUCTION • 8 MIN • PUBLISHED 18 SEPTEMBER 2026

Promoting a practice to an organisation-level belief is a counting argument over independent scopes — and one byte of punctuation stopped two repositories agreeing.

§1A convention is published before anyone asks for it9

§2A counting argument can be re-run; a judgement cannot10

§3The defect: a full stop made one word into two11

§4What a count cannot establish, stated before anyone asks12

References13

003

What a project that does not exist yet may be *told*

LIMITS • 8 MIN • PUBLISHED 18 SEPTEMBER 2026

Evidence that travels with the claim, a default that says in its own words it may be overridden, and a boundary a count cannot cross.

§1A first day arrives as 435 characters14

§2A default that says in its own words it may be overridden15

§3The line around a repository is a structure, not a filter15

§4Where a count of scopes stops being interpretable17

References18

END

Colophon

SOURCES AND THEIR HASHES, LICENCE, AND HOW THIS DOCUMENT IS SET

20

Rarity is a property of the *corpus*.

A stage whose job is to decide what an organisation believes returned two sentences, each citing as many as 41 repositories as having independently arrived at it. None of them had, and nothing raised an error. The weighting that decides which words in a sentence carry its meaning is **a statement about a collection, not about the sentence** — and the wrong collection runs the measure backwards where the data is most realistic.

PAPER • 001 PUBLISHED • 18 SEP 2026 READING • 8 MIN

LICENCE • CC BY 4.0

Canonical • digital1.foundation/articles/the-inherited-default/001-rarity-is-a-property-of-the-corpus.html

§1

The failure arrived as a well-formed answer

An organisation's engineering knowledge is written down inside individual repositories, where it stays. A practice several teams arrived at independently is the organisation's, and nothing in the store says so. The stage these notes are about promotes such a practice to an organisation-level statement once it appears in two or more distinct scopes, and what it promotes is later handed to projects that did not exist when any of it was written. It is counting and string comparison.

Re-measured against a corpus built to have the shape an engineering corpus actually has, it returned two groups where twenty conventions had been planted, the larger holding 41 members.^[7] A group elects one sentence and publishes it as something the organisation believes, cited to every scope inside it. Nothing in that result was catchable: no exception, no empty set, no value out of range.

THE PREMISE, PRECISELY

The stage asks two questions that look alike and are not. ***Which records might be about the same practice?*** is about the whole store; ***do these three say the same thing?*** is about those three. One weighting was answering both.

§2

A planted corpus, because the real one is private

A claim about a measure needs a corpus the reader can hold too, and the store this was developed against is private. So the experiment was planted: twenty conventions, each written three times, each of the three in a different scope, all twenty sharing one frame and differing in a single word. Around them sit ordinary notes, one per scope, in two shapes. The realistic shape writes those notes about the same nouns the conventions are about — an organisation writes about its migrations constantly and writes the convention down once. The control shares no vocabulary with them.

```
// 20 conventions × 3 phrasings, each phrasing in its own scope
const SUBJECTS = [
  "migration", "endpoint", "dashboard", "scheduler", "importer",
  "webhook", "renderer", "validator", "exporter", "indexer",
  "gateway", "collector", "publisher", "resolver", "uploader",
  "tokenizer", "planner", "archiver", "notifier", "compiler"
];

const PHRASINGS = [
  (s) => `Every ${s} is reviewed and tested before it reaches production.`,
  (s) => `A ${s} must be reviewed and tested before reaching production.`,
  (s) => `${s}s are reviewed and tested before they reach production.`
];

const corpus = [];
for (const s of SUBJECTS)
  PHRASINGS.forEach((p, i) => corpus.push({ text: p(s), scope: `acme/${s}-service-${i}` }));

// one ordinary note per scope, to give the pool a realistic size
const NOISE = Number(process.argv[2] ?? 3000); // 0 - 100 - 560 - 3000
const MODE = process.argv[3] ?? "subjects"; // realistic | control
const OTHER = ["cache", "buffer", "socket", "cursor", "lease",
  "digest", "shard", "replica", "quota", "probe"];

for (let i = 0; i < NOISE; i++) {
  const noun = MODE === "unrelated" ? OTHER[i % 10] : SUBJECTS[i % 20];
  corpus.push({ text: `Note ${i} about ${noun} config value ${i} in build ${i * 7}.`,
    scope: `acme/noise-${i}` });
}

const SIM = 0.55; // the bar every arm is run at
```

Two sentences count as the same practice when their content tokens overlap enough — shared tokens over the union of them, against a configured bar of 0.55. One flat number cannot do that job. Unweighted, pairs from inside one convention overlap 0.667 to 1.000, mean 0.778; pairs from two different conventions, 0.429 to 0.667, mean 0.561. The ranges meet at 0.667, and that overlap is the whole reason a weighting exists. ^[6]

Neither piece of machinery is this paper's. Weighting a token by how few documents contain it is term specificity, stated in 1972; ^[1] single linkage's habit of extending a chain without regard to the shape of the cluster it builds is textbook, and named — "this chaining can occur even when valid clusters exist". ^{[3][4]} Both behave exactly as their literature says. The defect is in the collection the weights were computed over.

Recovery tracks one ratio and nothing else

Take one planted convention: *every migration is reviewed and tested before it reaches production*. `migration` says which convention this is; `production` belongs to the frame all twenty share, so for the grouping to hold the first must outweigh the second. Weigh them over the whole candidate pool and the opposite happens as the pool grows, precisely because the corpus is realistic: the subject is common in an organisation's memory and the frame is rare. Pool-wide rarity measures how often a subject is discussed — not all equally frequent words are equally meaningful. [5]

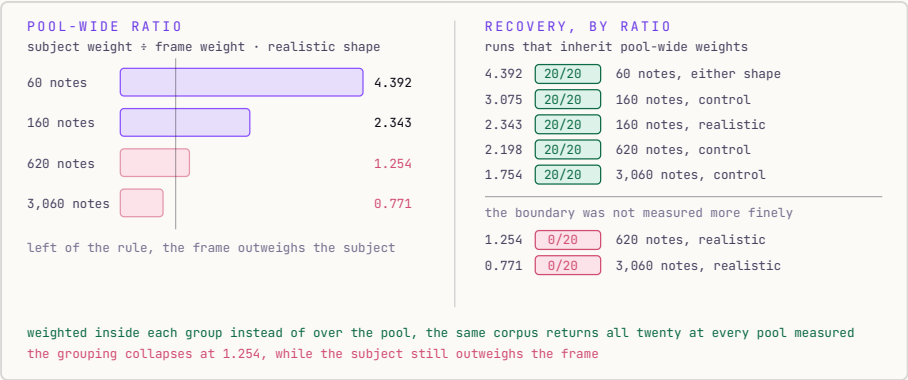


FIGURE 1 Recovery is a property of the corpus, not of the sentences. Bars are subject weight over frame weight; pills are how many conventions came back as their own three members.

Four arms were run over each corpus: single linkage unweighted; with pool-wide weights; the same with each group re-formed around a representative, which is the arm that was shipping; and the same with each group judged against weights computed over that group alone. **Recovered** counts conventions returned as their own three members and nothing else. The table is the realistic shape; the control runs are in Figure 1. [7]

RECORDS	RATIO	ARM	RECOVERED	CONTAMINATED	LARGEST	MS
60	4.392	all four arms	20/20	0	3	1–3
160	2.343	single linkage	0/20	1	60	13
160	2.343	other three arms	20/20	0	5	15–18
620	1.254	single linkage	0/20	1	112	53
620	1.254	+ idf	0/20	1	60	45
620	1.254	+ idf + refine	0/20	2	41	41
620	1.254	+ local idf	20/20	0	3	57
3,060	0.771	single linkage	0/20	1	600	647
3,060	0.771	+ idf	0/20	1	60	795
3,060	0.771	+ idf + refine	0/20	2	41	748
3,060	0.771	+ local idf	20/20	0	3	920

DEFECT · FOUND BY RE-MEASURING AGAINST A PLANTED CORPUS, 18 SEPTEMBER 2026

Words shared by the whole planted set outweigh the word that tells its members apart

Nothing in the three records below changes between the runs. The pool they are weighted against does, and with it the answer.

```
three records, three scopes – acme/migration-service-0, -1, -2
Every migration is reviewed and tested before it reaches production.
A migration must be reviewed and tested before reaching production.
migrations are reviewed and tested before they reach production.

pool of 60 · idf migration 1.166 · production 0.265 · ratio 4.392
one group of 3 · 20 of 20 conventions recovered, by each of the four arms

pool of 3,060 · idf migration 0.383 · production 0.497 · ratio 0.771
frame outweighs subject
0 of 20 recovered · 2 groups mixing conventions · largest 41 · 748ms
(+ idf + refine, the shipping arm)
```

Unweighted single linkage chains 600 members into one group – a blob nobody would publish, because nothing that size survives a look. The weighted arms fail quietly instead: two groups, the larger 41 members, each looking exactly like a practice held across the organisation.

§4

A group judged against itself

The correction leaves the find pass alone: pool-wide weights still drive it, because its job there is recall and over-merging is undone by the step after it. What changes is that the refinement stops inheriting those

weights and judges each candidate group against itself.

```
const idf = idfOf(candidates); // find: pool-wide
const grouped = groupBySimilarity(candidates, similarity, idf)
  .flatMap((g) => refineGroup(g, similarity)); // judge: group-local
```

Inside a group of three statements of one practice the frame is in every member and carries no weight; the subject is what remains. That is the `local idf` arm, and it recovers the twenty at each of the eight corpora measured, including the two where the pool-wide ratio has fallen below the boundary. It costs little — 920ms against 748ms at 3,060 records, 10.3s against 10.1s at 10,060 — and it does not fire eagerly: run against the store these notes were developed on, it promoted one convention, cited to three scopes, out of 326 distilled facts.

§ 5

What the planted corpus cannot settle

A RATIO THE RIGHT WAY UP IS NOT HEALTH

The obvious check does not work. Recovery is already 0/20 at 620 records in the realistic shape, where the ratio is 1.254 and the subject still outweighs the frame; outright inversion arrives only at 3,060, at 0.771. So the grouping stops working while the weighting still points the right way, and a monitor waiting for that ratio to cross 1.0 reports health through the whole failure it was written to catch. Where the boundary sits between 1.754 and 1.254 was not measured more finely, and the ratio is not offered as a tuning value.

The second limit is cost: over the range measured the find pass does not scale in proportion to the records, rising from 647ms at 3,060 to 10,129ms at 10,060 in the realistic shape — 3.3× the records for 15.7× the time. The third is the threshold: 0.55 is a configured default, never tuned against labelled data, and the unweighted ranges in §2 overlap.

The fourth is provenance for the claim itself. That rareness does not imply discriminating power was measured in 1975: terms occurring in one, two or three documents had "rather poor average discrimination ranks".^[2] What is added here is the mirror case, where the *frequent* token is the discriminating one — which arises because this collection is one organisation's repositories, not a general corpus. We looked for a treatment of that case and did not find one, so it is offered as an observation from one planted corpus, not a result standing on prior work.

ON THE PROVENANCE OF THIS MATERIAL

The method and measurements here were developed inside **SkilledMind**, a Digital One property; the stage described is its convergence pass. That repository is not public, so the engine these notes draw on — the grouping functions, the tests pinning the fix — is its builders' account, not something a reader can check. What a reader can check is the planted corpus: its generator is printed in §2, and the runs in §3 reproduce against any implementation of token overlap and single linkage. The two store figures — one convention, 326 distilled facts — come from that private store. Deliberately absent: prices, plans and tiers, and deployment internals.

References

- 1 K. Spärck Jones, *A statistical interpretation of term specificity and its application in retrieval*, Journal of Documentation 28(1), 1972, pp. 11–21; DOI 10.1108/eb026526, paywalled. Read from the 2004 reprint, staff.city.ac.uk/~sbrp622/idfpapers/ksj_orig.pdf, 18 September 2026.
- 2 G. Salton, A. Wong, C. S. Yang, *A Vector Space Model for Automatic Indexing*, CACM 18(11), 1975, pp. 613–620, discrimination-value section; printed page of the quoted passage unconfirmed. The same figure finds very high-frequency terms worse discriminators still. openlib.org/home/krichel/courses/lis618/readings/salton75.pdf, 18 September 2026.
- 3 A. K. Jain, R. C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, 1988, printed p. 75 = PDF p. 89; no text layer, read as page images. homepages.inf.ed.ac.uk/rbf/BOOKS/JAIN/Clustering_Jain_Dubes.pdf, 18 September 2026.
- 4 C. D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008, §17.2 — freely readable, same property. nlp.stanford.edu/IR-book/html/htmledition/single-link-and-complete-link-clustering-1.html, 18 September 2026.
- 5 K. W. Church, W. A. Gale, *Inverse Document Frequency (IDF): A Measure of Deviations from Poisson*, Third Workshop on Very Large Corpora, 1995, abstract. aclanthology.org/W95-0110, 18 September 2026.
- 6 O. Chum, J. Philbin, A. Zisserman, *Near Duplicate Image Detection: min-Hash and tf-idf Weighting*, BMVC 2008, for rarity-weighted set resemblance; A. Z. Broder, *On the resemblance and containment of documents*, SEQUENCES 1997. The coefficient's own 1901 and 1912 papers could not be opened and are not quoted.
- 7 The harness: one file building the corpus printed in §2 and running the four arms. Its grouping functions are not public, so the generator is printed instead. One workstation, 18 September 2026. Not linked.

A convention is a count, not a *judgement*.

A practice several teams arrived at separately is the organisation's, and something has to decide when that has happened. These notes derive the decision from the shape of its output — one sentence, published to every project in the organisation, carrying a count of distinct scopes as the whole of its warrant — which forces **two independent scopes and no model in the loop**. Those properties rest on a comparison of two strings, and **one byte of punctuation broke it**.

PAPER · 002 PUBLISHED · 18 SEP 2026 READING · 8 MIN

LICENCE · CC BY 4.0

Canonical · digital1.foundation/articles/the-inherited-default/002-a-convention-is-a-count.html

§1

A convention is published before anyone asks for it

Start from the output. The promotion stage produces no score and no ranking. It produces one sentence, in the organisation's own words, addressed to every project inside it and asserted as a thing the organisation does — carrying, as the whole of its warrant, the count of distinct scopes the practice was found in. ^[1]

Nothing about it is negotiated afterwards. A project that did not exist when any of the material was written is handed the organisation scope and nothing else, so this promotion is the only route by which one repository's practice can reach it. ^[2] The sentence arrives already believed.

One requirement falls straight out of that. A scope marked **sensitive** contributes to no convention, because a convention's destination is the whole organisation. ^[1] A promotion that could be talked into reading a sensitive scope would not leak a candidate — the persuasion would itself be the publication.

THE REQUIREMENT, PRECISELY

A practice is promoted only when statements of it are found in at least two distinct scopes. The unit of the count is the scope — a repository, a bounded place where one team writes things down — not the sentence, the file or the author: counting sentences would measure how thoroughly a team documents itself, counting scopes how many separate places arrived at the practice. A scope contributes once, however often it repeats itself.

It does not over-fire at the size measured. Run against the local store on 18 September 2026, the pass had promoted exactly one convention, cited to three scopes, among 326 distilled facts. ^[3]

§2

A counting argument can be re-run; a judgement cannot

There is an obvious alternative: hand the candidates to a model and ask whether they express one organisational convention. It is declined, not because the answer would be poor, but because a sentence published as something an organisation holds has to survive the question *why does this say that*. A count answers in public — these scopes, this overlap, this threshold, over this corpus — every term of it re-derivable by whoever holds the corpus. A judgement answers with a provider, a model version, a prompt and a date, none of which is in the corpus.

Three costs follow. The organisation's institutional layer would depend on a vendor's prompt, so a revision nobody inside the organisation saw could change what it is recorded as believing. It would cost a call for every candidate statement. And it would drift: two runs over an unchanged corpus can disagree, so what an organisation believes could move on a day when nothing in the organisation moved. ^[1] The overlap measure standing in its place is cruder and has the opposite properties — auditable, free, identical each run. That does not make it right; it makes it printable, and the next section prints one of its failures.

WHAT MUST NEVER BE TRADED AWAY

The sentence and its warrant travel together. A convention's warrant is the count and the corpus it was counted over; once that warrant moves into a model's weights the sentence keeps its authority and loses its evidence, and nothing on the page distinguishes those two states from outside.

Turn the construction over and the exposure is plain. Independence is a count of scopes — but only of the scopes whose statements the comparison matched — and declining the model is worth something only if that comparison is right about what two statements say. It reduces

each statement to a set of content tokens and scores the intersection over the union: set resemblance, in the form Broder reduced to "set intersection problems".^[4] A pair is one practice at 0.55 and above, a candidate runs 25 to 400 characters, and at most 1,200 candidates are considered — each number configured rather than derived.^[1]

§3

The defect: a full stop made one word into two

The token pattern is `/[a-z0-9][a-z0-9._/-]{1,}/g`, and the dot sits inside the character class on purpose: `crunch.md`, `package.json`, `feature/x-y` and `rate-limit` are single identifiers here.^[1] The cost was charged at the other end of the sentence. A word ending a sentence kept its stop, so `production.` and `production` were two different tokens, and the last content word of every sentence that ended in one matched nothing.

DEFECT · FOUND BY RE-MEASURING THE TOKEN COMPARISON, 18 SEPTEMBER 2026

A trailing full stop made the last word of a sentence a different word

Nothing exotic reproduces it: two statements of one practice, one per repository, identical to the byte except that one of them ends the way sentences end.

```
the failing input · one practice, written down in two scopes
scope 1  Every service ships with a health endpoint.
scope 2  Every service ships with a health endpoint
                                                ^ one character

token-set overlap on four such pairs · the bar a pair must clear is 0.55

Every service ships with a health endpoint.    before  after
Database access goes through the repository layer.  0.714  1.000
We deploy on Thursday.                        0.250  0.667
Secrets are read from the environment.          0.400  0.750

two of the four under the bar: held in two scopes, counted in one
strip a TRAILING run of [._/-] only, and crunch.md stays one token
```

For those two the organisation held a practice in two places and the count saw one; the other two survive with their margin eaten. The fix is one line — strip a trailing run of `[._/-]` and nothing else — pinned by a test named for the failure rather than the fix: that the comparison "does not let a full stop make the last word of a sentence a different word".^[5]

The arithmetic underneath is worth setting out. Of `Every service ships with a health endpoint.` the comparison sees four words — `service`, `ship`, `health`, `endpoint` — short tokens and a stop list never reaching it, a trailing plural folding to the singular. Two sets of four differing in one

member share three of the five words in their union, and three fifths is the 0.600 that was measured. The stop takes one word out of the intersection and puts a second, unmatchable word into the union, so an overlap of *i* over *u* arrives as *i*-1 over *u*+1. Two of the four do not reach 1.000 even after the strip, so those statements differ in more than their punctuation.



FIGURE 1 The bar is where two statements become one practice, so a trailing byte decides which side of it a pair lands on — and a pair landing short is a practice the organisation holds and the count does not see.

It never surfaced as an error, only as a smaller count, which is indistinguishable from an organisation that has not converged yet. A convention that is not minted leaves no trace, because the pass reports what it promoted, what it already knew and what it retired, never what it should have promoted and did not.

A second instance has the same shape and a different byte: three repositories held one convention and two converged, because one had written `services ship with a health endpoint` where the others wrote `service ships`. Folding trailing plurals recovered the third, and both fixes are pinned by tests named for the sentence they protect.^[5]

§ 4

What a count cannot establish, stated before anyone asks

The threshold is a configured default, never tuned against labelled data. Measured over a planted corpus before any grouping, unweighted overlap within one convention runs 0.667 to 1.000 and across different conventions 0.429 to 0.667.^[3] The ranges meet, so no single unweighted threshold separates them: 0.55 is not what makes the grouping work — the weighting above it is, which is the previous paper's subject.^[6]

The plural fold is blunt on purpose: trailing plurals, and neither tense nor derivation, with carve-outs so that `class`, `status` and `analysis` survive, and one collision it knowingly accepts — `https` folds to `http`.^[1] Neither the fold nor the comparison sees word order.

The largest limit is in the count itself. Two distinct scopes are two places, not two judgements: two repositories scaffolded by one team in one week are a single decision wearing two names, and nothing inside a count

of scopes can tell. The same gap leaves the sharing of conventions across an organisational boundary an open question rather than a feature: two sub-organisations sharing a platform team are not two independent observations either, and whatever settles that is about provenance rather than similarity. Not built, not measured.

None of it is re-runnable by a reader today: the engine's repository is not public, and the store figure in §1 was taken against a private store.^[3] What is printed here is enough to re-derive the arithmetic, not enough to reproduce the measurement. A count answers how many places wrote a practice down. It does not answer how many times that practice was decided.

ON THE PROVENANCE OF THIS MATERIAL

The stage described here was built inside *SkilledMind*, a Digital One property; it is the provenance of this method, not its subject. Its repository is not public, so what is quoted above is an account of that code rather than something a reader can open today — which is why the failing input and the four measured pairs are printed here in full. Both those and the store figure come from one workstation on 18 September 2026, and the store is private. Deliberately absent: prices, plans and tiers, and deployment internals.

References

- 1 The promotion stage — `packages/consolidation/src/converge.ts`, file header: the rules quoted in §1 and §2, and the defaults beside them — threshold 0.55, length 25–400 characters, `maxCandidates` 1200, the token pattern, the plural fold. Not public.
- 2 The birth-brief stage — `apps/skilledmind-api/src/genesis.ts` (`ORG_SCOPE`) and its test file, holding the counterfactual leak test read in §2. Same standing as [1].
- 3 The measurements of 18 September 2026, on one workstation: the four pairs in §3 before and after the fix, and the pairwise ranges in §4 over a planted corpus with plain unweighted overlap. The store figure in §1 is private and cannot be re-run by a reader.
- 4 A. Z. Broder, *On the resemblance and containment of documents*, SEQUENCES '97, IEEE, abstract: "to reduce these issues to set intersection problems". His resemblance is unweighted, and the measure is older — Jaccard's coefficient of community — but neither Jaccard paper could be opened for these notes, so nothing here quotes them. cs.princeton.edu/courses/archive/spring13/cos598C/broder97resemblance.pdf, read 18 September 2026.
- 5 The regression tests — `packages/consolidation/test/converge.test.ts`: the one named *does not let a full stop make the last word of a sentence a different word*, the assertion that `crunch.md` and `package.json` survive tokenisation whole, and the plural fold. Not public; the failing input is printed in §3.
- 6 The Inherited Default, Paper 001 — *Rarity is a property of the corpus*: the planted corpus whose generator it prints, and the weighting that answers the ranges quoted in §4.

What a project that does not exist yet may be *told*.

A repository can be handed what its organisation settled before that repository existed, on three conditions: **the evidence travels with the claim**, the text says in its own words that it may be overridden, and the line keeping one team's private specifics out is structural rather than a filter someone remembers to apply. Measured on 18 September 2026 it arrives as 435 characters, one recall, no model call. Across an organisational boundary the first condition stops being checkable — **a count of independent scopes assumes the scopes are independent** — and there this work stops.

PAPER • 003 PUBLISHED • 18 SEP 2026 READING • 8 MIN

LICENCE • CC BY 4.0

Canonical • digital1.foundation/articles/the-inherited-default/003-what-a-new-project-may-be-told.html

§1

A first day arrives as 435 characters

An organisation's tenth repository starts where its first did: a new checkout has no way to ask what the other nine settled. A birth brief is that ask: one recall against the organisation scope, no model call, 8.6 ms on the store measured here.^[1] On 18 September 2026 the answer was 435 characters — 68 words, one item, 79 characters of it the organisation's own sentence. The rest is the frame, and the frame decides how the sentence is read.

birth brief • one item • framing verbatim, re-wrapped; varying parts marked
How this organisation usually builds a service. ← the label tracks the kind

These are DEFAULTS, not rules — drawn from what the org already does, not from a policy. Follow them unless this project has a reason not to; when you deviate, say why, and that becomes the next thing worth remembering.

- [the org's own sentence • 79 characters • withheld] _(independently practised in 3 projects)_

Being able to ask is paid for once. Driven over stdio with a JSON-RPC client rather than read off the source, the bridge exposes five tools and returned a tool list of 3,616 bytes, 80 to 93 ms after the process started.

[2] Under revision 2026-07-28 of the Model Context Protocol – stateless and per-request, the `initialize` handshake older write-ups describe being legacy there – a server declaring the tools capability "MUST respond to `tools/list` requests with the set of tools currently available to the requesting client". [3] That is roughly 977 tokens standing in context for the session, against about 118 for the brief – estimates at 3.7 bytes per token, not vendor counts, and the payload itself varies by a few bytes of framing between drives.

§2

A default that says in its own words it may be overridden

An organisation that converged on one way of doing something should hand a new project that way as a starting point, not as a constraint: the project that needs the other thing is exactly the one a shared memory must not argue with. The mechanism is not configuration. It is the fixed middle paragraph printed above, carried by every brief that is not empty, and it is fixed because the reader is a model and the framing decides the outcome more reliably than the facts do. [1] Deviation is invited rather than tolerated, and a deviation with a reason attached is where the next convention comes from.

The evidence riding in the same line does the other half. `_(independently practised in 3 projects)_` is a number a reader can weigh and, inside the organisation, check against the scopes carrying it. Strip it and the sentence becomes an instruction from nowhere, which a model follows just as readily.

A FRAMING IS NOT AN ENFORCEMENT

What a test pins is that the words are present: that a rendered brief states these are defaults and invites deviation, and that a tenant with nothing distilled gets a brief saying so rather than an invented default. [4] No measurement here establishes how a particular model behaves on reading them. Nothing in this engine stops an agent treating the paragraph as policy, and nothing counts how often one does.

§3

The line around a repository is a structure, not a filter

The worst sentence a brief could carry is specific: one team's private operational detail, correct and relevant, handed to projects that have nothing to do with that team. The defence is not that such details are recognised and removed. The module assembling a brief hands recall a constant scope list – the organisation scope and nothing else, commented in the source as the security boundary – so a repository-scope memory is not something that code can reach. [1]

One route is therefore left out of a repository, and it is promotion. A practice becomes the organisation's when at least two *distinct* scopes recorded it independently — one repository repeating itself is a style — and a scope marked *sensitive* contributes to no convention at all, because a convention is published to everyone inside the organisation by construction.^[5] What that count does and does not establish is the previous paper.^[6]

An absence is not self-evidently a boundary holding: a brief that omits a memory because nothing matched it looks identical to one that omits it because it was never eligible. So the test asserts both halves:^[4]

```
planted · a repo-scope memory, specific and one team's business
scope   repo acme/private-web
text    "The web frontend deploys through the Falconridge pipeline
        with the staging bypass token"

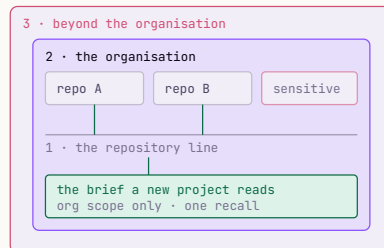
asked as a new webapp would ask
no "Falconridge" in the brief, and none in any item

the counterfactual, in that memory's own scope
hits contain "Falconridge" ← stored, indexed, retrievable by its owner
```

WHAT THE STRUCTURAL GUARANTEE DOES NOT SAY

It says this module reads one scope. It does not say a private thing can never reach a new project by another route: promotion is itself a route, and a specific that two teams wrote down separately is, by this design's logic, no longer one team's. Nor does it cover a store whose scopes were labelled wrongly, or what an operator holding database credentials can read.

THREE LINES, OUTWARD



lines 1 and 2 are decided by things a reader can point at: a count of scopes, and one constant scope list
 line 3 would read that count where nothing shows the scopes it counts are independent of one another

WHAT MAY CROSS

- 1 · out of a repository
 a practice, as one sentence
 on condition: 2 or more distinct scopes
 a sensitive scope contributes nothing
- 2 · within the organisation
 a convention in the org scope, ranked
 on condition: none — inside is inside
 which is why line 1 is gated so hard
- 3 · out of the organisation
 nothing: not built, not measured
 the condition would be independence,
 which a count cannot establish

FIGURE 1 Two of the three lines are decided by something a reader can point at; the third is a line this design has no rule for, so the drawing marks it.

Where a count of scopes stops being interpretable

Walk outward once more. A global organisation wants a practice settled in one sub-organisation available in the others; two organisations that trust each other want to exchange what they have distilled without exchanging repositories. Neither is built and neither is measured, and the reason is not effort: the number that carries a convention out of a repository stops being interpretable at that line.

OPEN PROBLEM • IDENTIFIED 18 SEPTEMBER 2026

A count of scopes cannot establish independence across an organisational boundary

Everything above rests on one number meaning one thing. A convention's authority is a count of **independent** scopes: not that a sentence appears twice, but that two parties arrived at it separately, which is why one repository repeating itself is excluded by rule. ^[5] Inside a single organisation the scopes are enumerated by the store that holds them, so distinctness and independence coincide closely enough to count on.

Across a boundary they come apart, and the count cannot tell. Two sub-organisations that share a platform team are not two independent observations: they are one practice, written down twice by people who talk to each other. Two organisations that arrived at it separately are two. The arithmetic is identical in both cases and returns the same number, so nothing in the result separates the reading that earns authority from the reading that borrows it — and the error runs in the direction that hurts: a convention resting on a doubled observation does not look weaker than one resting on two, and the count is what licenses a sentence to be published to everyone.

The shape of an answer is therefore not a better measure. Similarity is a property of the sentences, and no refinement of it reaches independence, which is a property of the people and the processes that wrote them: whether a practice was arrived at twice or copied once is nowhere in the words. Whatever settles this is a statement about **provenance** — where each observation came from, and separately from whom — travelling with the sentence and checkable by the party receiving it. What would have to be established is that two contributed scopes do not share the origin that produced the practice — an assertion about how an organisation is structured, not one anybody outside it can verify by inspecting text.

Nothing has been built for any of that: no exchange format, no provenance field, no measurement, no experiment that would bound the error. It is written down because it is the first thing a reader of the three conditions will ask, not because these notes answer it. **This block carries no failing input, because nothing is built to fail.**

The rest of what is open, with what would settle each. **The store measured holds one convention** among 326 distilled facts, so the kind-specific expansion had nothing to choose between and returned the same item for the webapp, api, service and infra kinds: the ranking behind §1 is untested at that size, and a store of tens of conventions would settle it. **The 0.55 similarity threshold** was never tuned against labelled data,^[5] and because it moves what an organisation is told it believes, the decision is joint rather than an engineer's. Nothing here is committed.

ON THE PROVENANCE OF THIS MATERIAL

The engine and the measurements come from *SkilledMind*, a Digital One property, which is the provenance of this method and not its subject. Its repository is not public, so the promotion rule, the brief's framing text and the boundary test are quoted from files a reader cannot fetch today, which is why they are printed above. The live figures were measured against a private store on one workstation and cannot be re-run by a reader; the organisation's own sentence inside the brief is withheld for the same reason. Token figures are an estimate at 3.7 bytes per token, not a vendor count. Deliberately absent: prices, plans and tiers, and deployment internals.

References

- 1 The birth-brief module — `apps/skilledmind-api/src/genesis.ts`: the constant organisation scope, commented there as the security boundary; the ranking rule; and the framing text printed verbatim in §1. Not public; live figures from a private local store, 18 September 2026.
- 2 The connection cost, measured by driving `packages/cli/sm-mcp.mjs` over stdio with a JSON-RPC client rather than by reading the source, 18 September 2026: five tools, a 3,616-byte `tools/list` payload, 80 to 93 ms to a listed tool set. A later drive returned a few bytes more, so the token figures beside them are estimates twice over. Not public.
- 3 *Model Context Protocol*, revision `2026-07-28`, Server Features — Tools. Cited for the cost measurement only: it settles how a tool list is served and says nothing about distilled organisational knowledge. modelcontextprotocol.io/specification/2026-07-28/server/tools, read 18 September 2026.

- 4 The leak test with its counterfactual, the defaults wording and the honest empty — `apps/skilledmind-api/test/genesis.test.ts`; the planted memory and both assertions are printed in §3. Not public; read 18 September 2026.
- 5 The promotion rules — `packages/consolidation/src/converge.ts`, file header: two or more distinct scopes, no `sensitive` scope, no model call; 25–400 characters and a threshold of 0.55, never tuned against labelled data. Not public.
- 6 The Inherited Default, Paper 002 — ***A convention is a count, not a judgement***: what a count of scopes establishes, and where the grouping behind it stops.

END

THE INHERITED DEFAULT · COLOPHON

Colophon

A

The record

This document is generated from the published pages, which are the record. Where this file and a page disagree, the page is right and this file is stale. Each source below is named with the SHA-256 of the file this document was built from, so a reader can check that what they are holding is what was published; the generator fetches each live page as it builds and reports any that has moved since.

INDEX	digital1.foundation/articles/the-inherited-default/
001	digital1.foundation/articles/the-inherited-default/001-rarity-is-a-property-of-the-corpus.html sha256 5d7d6a1bebf6b454b182932b249026a666829bee5f7f6cfe442e68b646f73213
002	digital1.foundation/articles/the-inherited-default/002-a-convention-is-a-count.html sha256 39c1563cd6e2d6df009db9251e732a894316b7053e66fc5161db6eff82e57123
003	digital1.foundation/articles/the-inherited-default/003-what-a-new-project-may-be-told.html sha256 abf6a6721d258093ce5a440d80da0a9ee9be8a79fb9c3745fcda4311548c8044

B

Licence

Every paper in this document is published under the Creative Commons Attribution 4.0 International licence, creativecommons.org/licenses/by/4.0/. You may copy, redistribute and adapt it for any purpose, including commercially, provided you credit **Stichting Digital One Foundation i.o.** and link to the page cited above.

C

How it is set

In the six faces the site itself carries and embeds: Fraunces 300 and 600 for titles and the emphasised word, Manrope 400 and 500 for the text, JetBrains Mono 400 for the ledger's own labels, code and running matter. Those six are Latin subsets and do not carry every character the papers use — ª is absent from them. Nothing is drawn in their place and quietly dropped from the text: each is set in a named system face. Every character in every paper was extracted back out of this file and checked before it was written, so the whole document copies, searches and cites as it reads.

Figures are the papers' own inline drawings, re-toned from the site's night ground to this page by a one-to-one swap of the closed house palette, so green still means it holds and rose still means it breaks — and the two are separated in weight of grey as well as in hue, so a photocopy keeps the distinction. The rail down the left of each page is the ledger from the Foundation's homepage; a filled dot marks each numbered section, and the § number sits in the gutter, outside the measure.

The sheet is A4 as the page asks for it, which Chromium rounds to 594.96 × 841.92 pt when it writes the file — 0.1 mm narrower and 0.03 mm taller than the 595.28 × 841.89 pt of the standard. No printer will show it; a document that prints its own hashes should say it anyway.

© MMXXVI Stichting Digital One Foundation i.o.
Amsterdam, the Netherlands

Published under CC BY 4.0.
No public-benefit (ANBI) status is claimed at this time.

digital1.foundation
The Inherited Default