Information Architecture Research: Fisheries and Oceans Canada

Improving findability through diagnostic tree testing

Nov 14, 2026 · @Ethan

Executive summary

In a 16-participant tree test, splitting one overloaded sub-category on the Fisheries and Oceans Canada (DFO) website raised success among participants who reached that section from 40% to 69%. Overall task success rose from 25% to 56%.

The problem. Participants were asked where they would find information on minimizing damage to undersea ecosystems. The correct page, *Protecting Aquatic Ecosystems*, sits two levels below Aquaculture, the category for fish farming, beside pages about fish disease. Four of 16 participants found it.

The diagnosis. My team's earlier course project treated this as a vocabulary problem and renamed Aquaculture to *Aquatic Farming*. Success fell. Reading that project's card sort and tree test together a year later pointed to a different cause. The page is correctly filed under Aquaculture, since it covers how fish farms limit their environmental impact. The problem was that nothing along the path to it signalled environmental content.

The intervention. I split the node holding the target in two. Fish-disease pages moved to a new *Fish Health Management* node; the ecosystem and reporting pages moved to *Environmental Impact and Reporting*. Nothing else in the tested tree changed.

The result. Overall success rose from 25% to 56% across two rounds of 16 participants, 15 of whom took part in both. Part of that gain came from more participants opening Aquaculture, which the revision did not cause. The cleaner measure is success among participants who opened Aquaculture: it rose from 40% to 69%, within one point of the figure I predicted before testing. Every participant who reached the new environmental node chose the correct page, up from four of five at baseline.

What it did not fix. Directness moved by one participant, and average task time rose from roughly 50 to 77 seconds. The revision made the destination recognizable; it did not make the route obvious.

At 16 participants, this study falls below the minimum of 30 commonly recommended for quantitative tree testing. Findings are reported as directional evidence supported by path analysis, not as statistically reliable effect sizes.

Background

This study extends my team's project from UX205 Information Design at Wilfrid Laurier University, which evaluated the information architecture of the Fisheries and Oceans Canada website.

DFO is the federal department responsible for Canada's fisheries, oceans and waterways. Its website serves commercial fishers, recreational anglers, Indigenous communities, industry operators, job seekers and the general public. Content spans licensing, regulations, conservation programs, scientific research and enforcement, a breadth that makes category boundaries hard to draw and discoverability hard to guarantee.

The original project

My team's project ran across three assignments, each building on the last.

Assignment	Method	Participants	What it established
1	Site inventory in Miro	None	The live structure: 7 top-level categories, 115 nodes across 4 levels
2	Card sort	9	How users group this content when left to themselves
3	Tree test, two rounds	8 per round	Whether users could find two target pages in the existing and revised structures

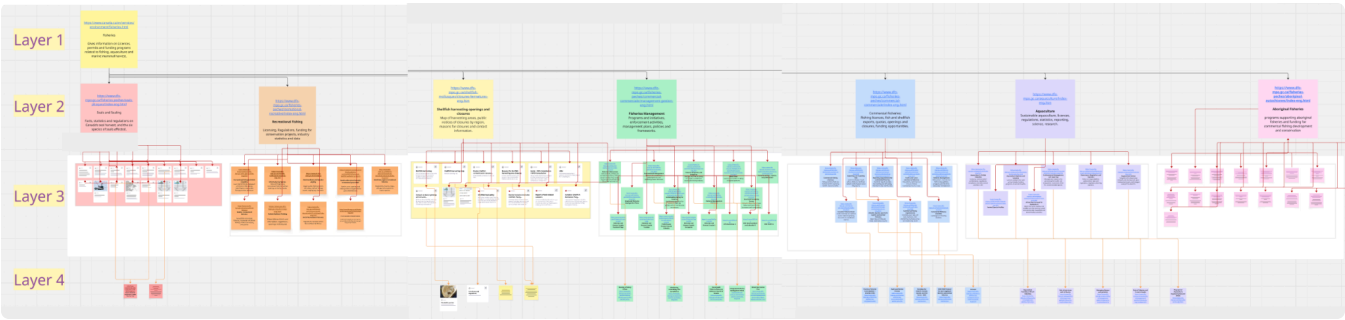

Figure 1. The site inventory built in Miro (Assignment 1): 115 nodes across four levels, colour-coded by top-level category. The complete tested tree appears in Appendix C.

The card sort found strong conceptual boundaries. Participants separated seal content, shellfish safety, commercial fishing, enforcement and aquaculture into distinct groups and rarely crossed between them. Aquaculture was the clearest case. Across all nine participants, the Aquaculture card was never placed in the same group as enforcement and management cards such as *Fisheries Enforcement* and *Become a Fisheries Officer*, commercial cards such as *Commercial Communal Fishing Licenses*, or any seal or shellfish card.

The tree test used two tasks. One asked where a user would find information on minimizing damage to undersea ecosystems; the other asked where to find fisheries enforcement jobs. Round one tested the live structure. Round two tested a revision with two changes: Aquaculture was renamed *Aquatic Farming*, and Fisheries Management was removed from the top level, leaving it reachable only through Commercial Fisheries.

Both changes reduced success. The ecosystems task fell from 6 of 8 participants to 3 of 8. The enforcement task fell from 6 of 8 to 5 of 8, and successful participants started from three different categories instead of converging on one.

Contribution and scope

Assignments 1 to 3 were produced by a team of 5. My own contributions were the similarity matrix, dendrogram and user-suggested-labels analysis for the card sort, and the pathing and revised-structure analysis for the tree test, alongside team brainstorming on test design.

Writing that analysis left my team with a result but not an explanation: we knew the revision had failed, not why. **A year later, I returned to the project on my own to find out, and to test whether I could design a structure that worked.** Everything from the re-diagnosis onward is my own work: the revised structure, both rounds of testing, recruitment, analysis, and this report.

Diagnosis

In hindsight, the original revision aimed at the wrong part of the tree. It changed a top-level word, when the harder problem sat inside Aquaculture: the environmental page was correctly filed but poorly signposted.

What the task asked

The task read: *"You are a marine animal safety activist, and you are looking to find information on minimizing damage to undersea ecosystems. Where would you find this?"*

Tree test participants navigate by matching the words in their goal to the labels in front of them. This task's key words are *marine*, *damage* and *ecosystems*. No top-level category shares any of them. Choosing Aquaculture requires a participant to already know that fish farming affects marine ecosystems. Nothing in the label states that connection, so participants have to infer it. The original paths show them hesitating: opening Aquaculture, backing out to Fisheries Management or Shellfish, then returning.

At the time, renaming Aquaculture was a reasonable response to what my team saw. Participants were choosing Fisheries Management over Aquaculture, so we concluded they did not know what *aquaculture* meant and renamed it *Aquatic Farming*. I was part of that decision. Reflecting on it a year later, with the card sort beside the tree test, a different explanation fits better. The new name made the category read even more clearly as fish production, which gave a participant looking for ecosystem content one more reason to look elsewhere.

What the card sort showed, read again

Reading the card sort again, part of the explanation was already in my team's own data. None of the nine card sort participants grouped the Aquaculture card with enforcement, management, commercial fishing, seal or shellfish cards. In their sorting, aquaculture stood apart from almost every other topic on the site.

That matters for navigation. If users see aquaculture as narrowly about producing fish, they are unlikely to expect environmental content inside it, even when that content belongs there. A clearer word for fish farming cannot change that expectation. What can help is making the environmental content easy to recognize once a user is inside.

What the site inventory showed

The site inventory shows where participants struggled once inside. *Protecting Aquatic Ecosystems* sits two levels below Aquaculture: Aquaculture → Environmental Managing and Reporting → Protecting Aquatic Ecosystems. The page belongs there. DFO's own summary of the page reads "Fish farm site selection, managing organic waste, escape prevention": in other words, how fish farms limit their effect on surrounding waters (Fisheries and Oceans Canada, n.d.-b).

Its siblings under that node were *Reducing Disease Risks*, *Managing Diseases and Parasites*, *Treating Outbreaks* and *Aquaculture Public Reporting*. Three of the five pages concern fish disease, and one reports on the aquaculture industry's performance. The one ecosystem page sat among them under a parent label that reads as record-keeping.

So the target was in the right category but poorly signposted: its parent label did not signal environmental content, and its neighbours were mostly about fish disease.

The conclusion

Three sources point the same way. My team's tree test ruled out a vocabulary-only explanation: renaming the top-level category made results worse. The card sort helps explain the hesitation: participants kept aquaculture apart from almost every other topic. The inventory showed where the remaining problem sits: inside Aquaculture, the environmental page is surrounded by fish-disease pages under a label that reads as record-keeping.

The problem, then, is neither that the page is in the wrong place nor that it is too deep. A correctly filed page two levels down can still be found when each level confirms the user is on the right track. Inside Aquaculture, nothing gave that confirmation.

Research question

Does separating ecosystem-protection content from fish-health content, within the category where it already lives, improve findability for a user seeking information on reducing harm to marine ecosystems?

Methodology

I ran two rounds of tree testing on ProvenByUsers: a baseline on the live structure, then the revised structure. Fifteen participants took part in both rounds.

Why tree testing

Tree testing presents the navigation hierarchy as text alone, with no visual design, search bar or homepage promotion. If a participant cannot find a page under those conditions, the cause lies in the structure or the labels. Since the diagnosis concerned how content is grouped and labelled, the method needed to remove everything that might compensate for weak labels.

Design

I used a within-subjects design, so that most participants completed both rounds and served as their own control. This removes the largest confound in small comparisons: a change in success that comes from testing different people, when the point is to test a different tree. The match is close but not exact. Fifteen participants completed both rounds; one completed only the baseline and another only the revised round, so each round still had 16 participants.

The baseline ran 16 October to 4 November 2026. It ran longer than planned because I recruited additional participants partway through to strengthen the sample. The revised round ran 6 to 10 November 2026.

Participants

I recruited participants through personal and university networks. The two participants who completed only one round explain why the group counts differ slightly: the revised-only participant was a UX student, and the baseline-only participant was a university student outside the program.

Group	Baseline	Revised	Site familiarity
UX students from the original course cohort	6	7	Studied this website for a full term
University students outside the program	8	7	None
Non-students	2	2	None

Student participants ranged from 19 to 26 years old. I recorded which participants had studied the site because my team's card sort had found that topic familiarity changes how people sort this content; tracking it let me ask the same question about navigation.

The task

One task was tested in both rounds, worded identically:

> "You are a marine animal safety activist, and you are looking to find information on minimizing damage to undersea ecosystems. Where would you find this?"

The correct destination was *Protecting Aquatic Ecosystems*.

I excluded my team's second task, locating fishery officer jobs. It had performed well in the original round one and sits in a branch this revision does not touch, so it would have measured a part of the tree that was not under investigation. Excluding it kept sessions short and participation high, at the cost of a control task. That trade-off is revisited under Limitations.

The brief

Participants received a written brief orienting them to the site: who DFO is, what the website covers, how the tree works, and an instruction to follow their first instinct. The full text appears in the appendix.

I included short definitions of two terms, *fisheries* and *aquaculture*. This departed from my team's original procedure and needs justification, since defining a term that appears in the tree can make a task easier.

The reasoning was comparability. Every participant in the original study had spent a term studying this website and knew both terms. Of my participants, only the UX students had studied the site, a year earlier, and five participants told me before testing that they did not know what the terms meant. Without definitions, many would have been guessing at two terms central to the structure. The definitions were meant to bring participants closer to the knowledge the original participants already had.

Procedure

Participants completed the test remotely and unmoderated. ProvenByUsers caps a single test at ten participants, so I built each round as two identical tests and combined the results. For the revised round, I sent each participant the brief again so they did not have to scroll back to find it.

Measures

In this report, a *top-level category* is one of the seven options on the tree's first level, a *sub-category* is the level below it (such as *Environmental Managing and Reporting*), and a *node* is any item in the tree. ProvenByUsers recorded four things for each participant:

- **Success:** whether the participant chose *Protecting Aquatic Ecosystems* as their final answer.

- **Directness:** whether they reached their final answer without backtracking. A direct success went straight to the target.
- **Time:** seconds from starting the task to choosing an answer.
- **Path:** every node the participant opened, in order.

From the paths I counted how many participants opened Aquaculture, how many reached the correct sub-category, and where each failure ended. I also derived **success once inside Aquaculture**: of the participants who opened Aquaculture at any point, the share who succeeded. The revision only changed nodes inside Aquaculture, so this is the measure that isolates its effect.

Data handling

Before analysis, I excluded entries with no recorded answers and duplicate attempts. One participant began the baseline three times; only the final, completed attempt is counted, and the earlier exposures are noted under Limitations. Two other baseline entries recorded no answers and were excluded; one belonged to the participant who later completed only the revised round. One participant's revised response was recorded twice, and I kept the second.

I matched individual responses to participants by response order and timestamp, and checked the order against individual results. One baseline result set collapses two identical paths into a single row, so those two participants cannot be told apart. This affects no group total: both were university students outside the program, and both failed on the same path.

Reading the numbers

With 16 participants per round, each person is worth about 6 percentage points. That shapes how every number in this report should be read.

- **Differences of one or two participants**, such as two direct successes against one, or three failures at a page falling to two, are within ordinary variation. They are reported as observations, not as evidence of an effect.
- **Larger differences**, such as overall success rising from 4 participants to 9, are stronger but still directional. Tree testing guidance recommends at least 30 participants, and 50 or more per tree when comparing structures; this study has 16.
- **Averages** are sensitive at this size. One revised-round participant took a path of more than 40 steps, which raises that round's average time noticeably.

What carries more weight than any single percentage is whether the paths changed in the way the prediction said they would.

Baseline findings

Four of 16 participants found *Protecting Aquatic Ecosystems* in the live structure, a 25% success rate. One reached it directly. Average time was 50 seconds.

For context, Jeff Sauro of MeasuringU found an average completion rate of 66% across 77 tree-test tasks from three studies, a figure he described as tentative (Sauro, 2013). Success rates vary a lot by task, so it is only a rough reference point. Even so, 25% sits well below it.

Where participants were lost

Ten of 16 participants opened Aquaculture. Five of those reached *Environmental Managing and Reporting*. Four selected the correct page.

So the structure loses people at every stage, but unevenly. The largest single drop is inside Aquaculture: half the participants who correctly identified the right top-level category then failed to identify the right sub-category. Once participants reached the correct sub-category, four of five chose correctly, so the final step was not where most were lost.

Failure destinations

Final destination	Participants
Aquaculture environmental risk assessments	3
Integrated fisheries management plans	2
Fishing regulations and requirements	2
Shellfish harvesting map	2
Reducing Disease Risks	1
Provincial and territorial fisheries	1
Contact our offices	1

Two patterns stand out.

The decoy inside Aquaculture. The most common single failure was *Aquaculture environmental risk assessments*, reached through *Aquaculture Science*. Three participants ended there. The page carries the word *environmental*, sits in the correct top-level category, and sounds like an assessment of harm. At the point of choice, it competed directly with *Environmental Managing and Reporting* and won.

The pull toward Fisheries Management. Seven of 16 participants opened Fisheries Management at some point, and four of the 12 failures ended in its pages. The category describes itself as covering programs, enforcement, policies and frameworks, which is a reasonable place to expect government content about protecting ecosystems. Participants were not navigating carelessly; they were applying a sound assumption to a structure that does not reward it.

Directness and time

Only one participant of 16 reached the target without backtracking. Several correct answers came after extended wandering: one participant passed through six different top-level categories before arriving. Low directness alongside moderate completion time indicates a structure that can eventually be solved by exhaustive search, but gives almost no one a confident route.

Hypothesis and prediction

Hypothesis. Six participants opened Aquaculture at baseline and still failed. Two left it for another category. The other four failed in ways the structure inside Aquaculture could explain: three chose *Aquaculture Science* over *Environmental Managing and Reporting*, because neither label clearly signals environmental content, and one reached the correct sub-category but chose a fish-disease page. Splitting the node addresses both. It gives the environmental content a label that signals it at the point of choice, and it shortens the final list: a participant who opens the environmental node sees two options instead of five, with the disease pages moved to a node of their own.

Stated before testing

I recorded three commitments before building the revised tree, so that the results could not be rationalized afterwards.

Mechanism. The split works inside Aquaculture, so it should only help participants who open Aquaculture. It should not change how many participants open Aquaculture in the first place, since nothing above that level changed.

Primary prediction. Among participants who open Aquaculture, success should rise from 40% to about 70%. I derived this figure from the baseline paths. Ten participants opened Aquaculture at baseline: four succeeded and six failed. I sorted the six failures by where they went wrong.

Where the failure happened	Participants	Could the split reach it?
Chose *Aquaculture Science*, ended at its environmental risk assessments page	3	Yes. The new label competes directly with *Aquaculture Science* at the same choice.
Reached the correct node, then chose a fish-disease page	1	Yes. The disease pages move to a node of their own.
Opened Aquaculture, then left it for another category	2	No. The change only affects choices made inside Aquaculture.

That leaves four failures the split could reach. A clearer label will not win over every participant, so I assumed most but not all would be recovered: three of the four. Four existing successes plus three recovered gives seven of ten, or 70%.

With 16 participants, I treated this as a check on direction and mechanism, not a precise forecast. If the split worked for the reasons I believed, conversion should rise toward 70%, and the recovered participants should come from the first two rows of the table.

Success thresholds. I also set two thresholds for the overall result. Changes in success are measured in percentage points: the plain difference between two percentages, so moving from 25% to 40% is a gain of 15 points. With 16 participants, each person is worth about 6 points, so 15 points is two or three people. I set 15 points as the minimum for a meaningful improvement, and 20 points as a stricter standard given how few people each point rests on. For directness, I set three or more direct successes against a baseline of one: if the new label made the route obvious, more participants should go straight from Aquaculture to the target without backtracking.

The second threshold matters as much as the first. Overall success can rise because participants wander more productively; directness rises only when the path itself becomes clearer. Setting both in advance meant I would have to report whichever result came in, including a miss.

The revision

I split one node in two. Nothing else in the tree changed.

Environmental Managing and Reporting held five pages: three about fish disease, one public reporting page for the aquaculture industry, and one about ecosystem protection. I separated them into two sibling nodes under Aquaculture.

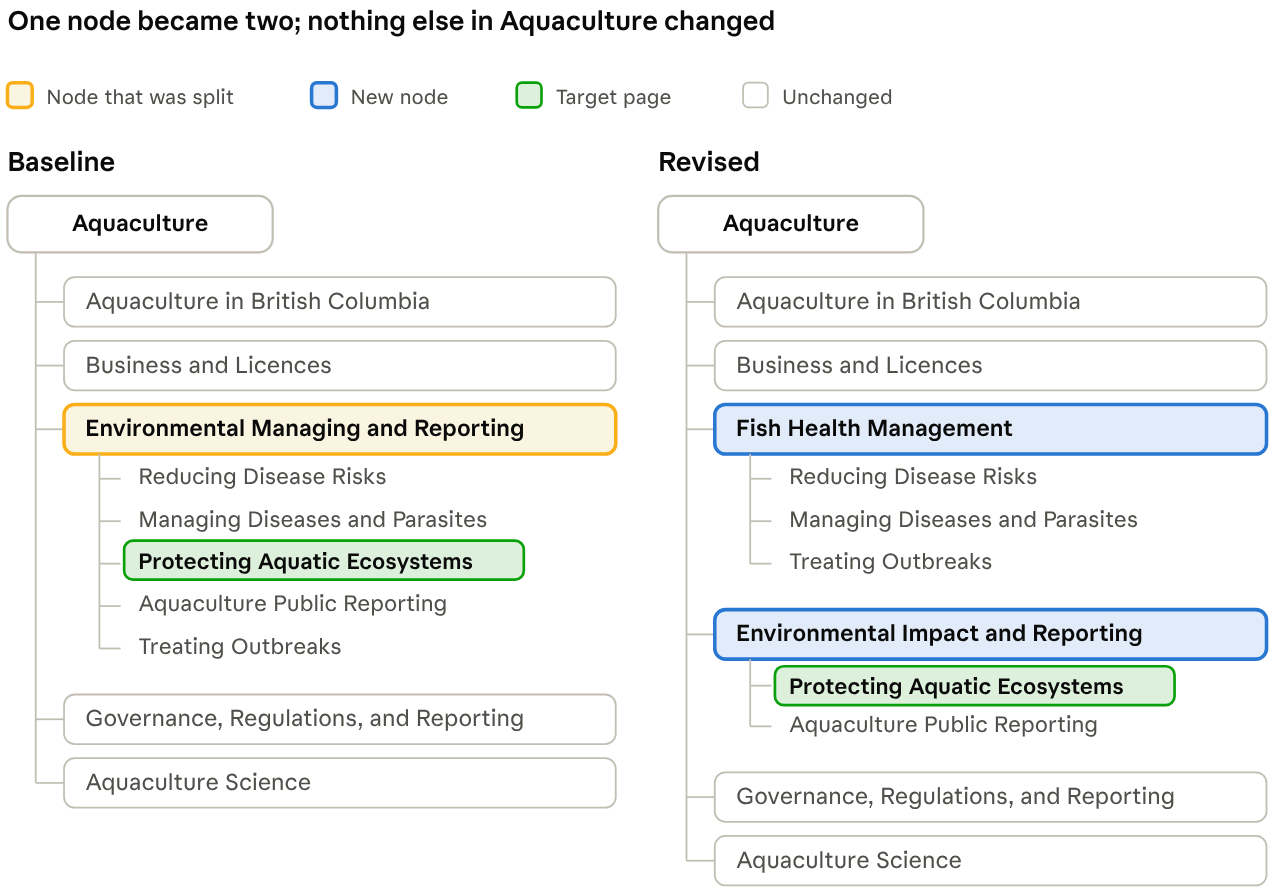

Figure 2. The Aquaculture branch of the tested tree, before and after the revision

The target stayed at the same depth, two levels below Aquaculture. It was not promoted, and its own label was not changed.

Why these labels

Fish Health Management describes its three pages literally and signals clearly that the node is about disease in farmed fish.

I chose *Environmental Impact and Reporting* over five alternatives, including *Ecosystem Impacts and Reporting* and *Aquatic Ecosystem Protection*. The word *impact* sits close to the task's *damage* without repeating the target page's own name, which would have made

the final step artificially easy. It also avoids *environmental protection*, a phrase with specific meaning in Canadian federal law that a government communications team would likely contest. And it describes both children honestly, since a reporting page sits under it alongside the ecosystem page.

What was deliberately not changed

I considered three other changes and rejected each. All three are defensible; I set them aside to keep the change narrow enough to interpret.

A new top-level environmental category. The card sort supports a separate environmental domain in principle. But when I reviewed the full 115-node site inventory, only two pages had clear ecosystem-protection content: *Protecting Aquatic Ecosystems* and *Ghost Gear*. Other candidates, such as an Indigenous habitat funding program, mention the environment but exist mainly to serve another purpose. A top-level category holding two pages, beside categories of up to 16, would look thin and do little for users.

Cross-linking the target into Fisheries Management. This was my first choice. Fisheries Management was the most common wrong turn: seven of 16 baseline participants opened it, and four of the 12 failures ended there. A cross-link would also have been the simplest change, adding one entry without moving anything. I set it aside for two reasons. Taxonomically it does not fit: Fisheries Management's pages are about managing the fishing industry, while *Protecting Aquatic Ecosystems* describes how fish farms limit their own environmental impact, which is aquaculture content. And once the baseline was complete, the larger leak was inside Aquaculture: half of the participants who opened it then failed to find the right sub-category. I record the cross-link as a candidate for future testing.

Removing or relabelling the decoy. *Aquaculture environmental risk assessments* drew three failures, but the name describes the page accurately. Stripping *environmental* from it would trade one findability problem for another.

Holding these changes back set a ceiling on what the revision could achieve. Because it changed only one node inside Aquaculture, it could only help participants who opened Aquaculture. At baseline, that covered six of the 12 failures. The other six participants never opened Aquaculture, and no change inside it could reach them. I knew this before testing.

Revised findings

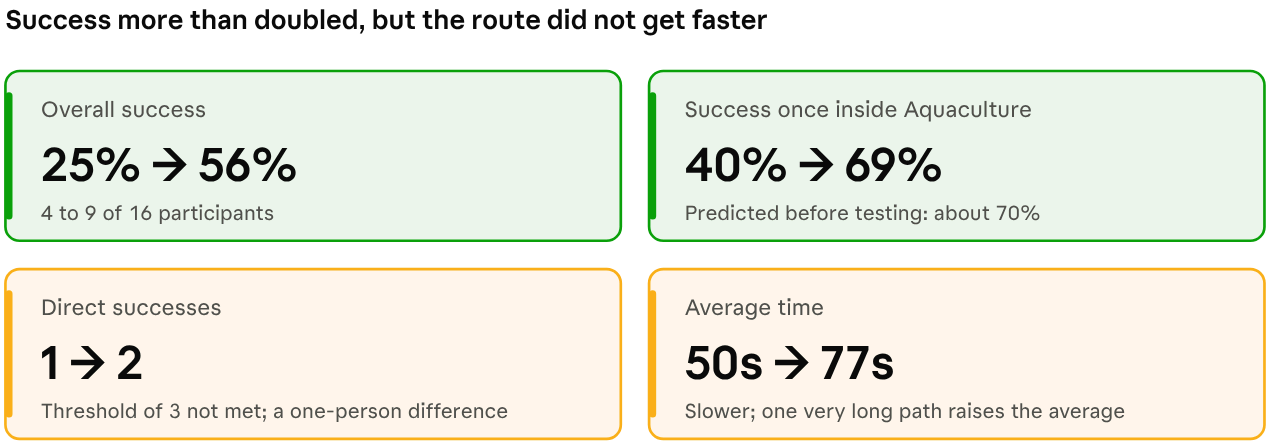

Figure 3. Key results, baseline to revised, 16 participants per round

The funnel

Thirteen of 16 participants opened Aquaculture. Nine reached *Environmental Impact and Reporting*. All nine selected the correct page.

The second and third figures are the result. At baseline, half the participants who opened Aquaculture failed to find the right sub-category; in the revised tree, nine of 13 found it. The final step lost nobody, against one participant at baseline, though a one-person difference is too small to read on its own.

Measured against the prediction

The prediction stated before testing was that success among Aquaculture entrants would rise from 40% to approximately 70%. The measured figure was 69.2%, nine of 13.

The exact match matters less than where the gains came from. The forecast said they should come from choices made inside Aquaculture, and they did: 9 of 13 participants who opened Aquaculture reached the correct sub-category, against 5 of 10 at baseline. Below that level, the counts are too small to read. Leakage to *Aquaculture Science* fell from three participants to two, and the disease-page failure at the correct sub-category did not recur. Each of those is a one-person difference, so they are consistent with the prediction but cannot confirm it.

Failure destinations

Final destination	Baseline	Revised
Aquaculture environmental risk assessments	3	2
Fishing regulations and requirements	2	0
Shellfish harvesting map	2	1
Integrated fisheries management plans	2	0
Policies and frameworks	0	2
Reducing Disease Risks	1	1
Contact our offices	1	1
Provincial and territorial fisheries	1	0

Two failure patterns remained, each resting on very few participants.

The decoy held. Two participants still reached *Aquaculture environmental risk assessments* through *Aquaculture Science*, against three at baseline. The new label won that comparison for most participants but not all. A change of one participant is within noise, but the page still catching people suggests the decoy is a separate problem from the one the split fixed.

The disease pull persisted, relocated. One participant went Aquaculture → *Fish Health Management* → *Reducing Disease Risks*. Reducing risks echoes minimizing damage, so the page attracts the same kind of reader. The split moved where this error happens, from the wrong page inside the right node to the wrong node entirely, without removing the pull. This rests on one participant in each round.

One destination appeared that had not at baseline: two participants ended at *Policies and frameworks* under Fisheries Management. Nothing in that branch changed, so this is most likely participant variation.

Comparative analysis

Measure	Baseline	Revised	Change
Participants	16	16	

Measure	Baseline	Revised	Change
Success	25.0% (4)	56.3% (9)	+31.3 pts
Direct success	1	2	+1
Opened Aquaculture	10 (62.5%)	13 (81.3%)	+18.8 pts
Reached correct sub-category	5	9	+4
Success once inside Aquaculture	40%	69.2%	+29.2 pts
Success once at correct sub-category	80%	100%	+20 pts
Average time	50s	77s	+27s

Green: improvement in the part of the tree the revision changed. Yellow: threshold missed, or a result that got worse. Grey: a change the revision cannot explain.

Against the thresholds I set in advance: The 15-point threshold for overall success was exceeded, as was the more conservative 20-point standard. The directness threshold of three was not met; two participants reached the target directly, against a baseline of one.

Which number to trust

The headline figure overstates the effect of the revision, and the reason is visible in the table.

Entry into Aquaculture rose from 10 participants to 13. Nothing above the Aquaculture level was altered between rounds: the seven top-level categories were identical in name, order and content. The revision cannot have caused that increase. It reflects some combination of participants remembering the task from the first round, individual variation, and chance in a sample of 16.

Part of the overall gain from 25% to 56% therefore comes from more participants arriving at the right category, which the revision did not cause.

The cleaner measure is **success among participants who opened Aquaculture**: 40% to 69.2%. It counts only people who reached the category, so it isolates the part of the tree that actually changed, and it is the figure I predicted in advance. I report both rates, but the conclusions rest on this one.

By participant group

Group	Baseline	Revised
UX students from the original cohort	1 of 6 (17%)	4 of 7 (57%)
University students outside the program	3 of 8 (38%)	4 of 7 (57%)
Non-students	0 of 2	1 of 2

The pattern runs against the obvious expectation. Participants who had studied this website for a term did worse than the other students at baseline, and improved most under the revision. Familiarity with the subject did not help them find ecosystem content inside a fish-farming category.

This is consistent with the original card sort, which found that participants with moderate topic knowledge sorted most consistently, while those with the most knowledge sorted too specifically. Prior exposure appears to shape expectations about where content *should* live, which the live structure then violates.

Two cautions. Group sizes of six to eight do not support claims about population differences, and the two rounds are not composed of identical groups: one participant differs in each direction. The pattern is worth noting and not worth concluding from.

Directness and time

Success rose by 31 points while directness moved by one participant and average time grew by more than half.

That combination has a specific meaning. If the revision had made the route clearer, directness would have risen and time would have fallen. Instead, more participants arrived, and they took longer getting there. Several successful paths in the revised round ran through five or more top-level categories before reaching Aquaculture.

The revision made the destination recognizable once reached. It did not make the route to it any more obvious.

What the revision solved, and what it did not

The revision improved recognition and left wayfinding largely unchanged.

Recognition is whether a user, looking at a list of options, can tell which one holds what they want. **Wayfinding** is whether the structure gives them a route to that list in the first place. The two are separable, and a structural change can improve one without touching the other.

The evidence sits in three numbers. Success at the correct sub-category went from four of five to nine of nine, meaning once participants saw *Environmental Impact and Reporting*, every one of them recognized it. Success inside Aquaculture went from 40% to 69%, which fits the new label being easier to recognize than the old one. But directness moved by one participant and average time grew by 27 seconds, because nothing about the journey to Aquaculture changed.

In practice, this means a participant arriving with an ecosystem question still has no clear first move. The revision rewards them once they guess Aquaculture correctly; it does not help them guess it. These participants were role-playing an activist, so a real activist's first move may differ (see Participant composition under Limitations).

What remains unsolved

The decoy. *Aquaculture environmental risk assessments* still drew two failures. It sits under *Aquaculture Science*, competes on the word *environmental*, and catches users who are already in the correct top-level category. It ties with Policies and frameworks as the most common remaining failure destination, and it is the only one inside Aquaculture.

The disease pull. *Reducing Disease Risks* still attracts one participant whose task is about minimizing damage. Separating it into its own node changed where the error happens but not that it happens.

The top-level problem. Three participants in the revised round never entered Aquaculture at all, ending in Shellfish and Fisheries Management. Nothing in this revision addresses the question of why a user with an ecosystem question would look under a fish-farming heading in the first place.

That last point is the ceiling on this work. No top-level label on the DFO site uses the words a user with an ecosystem question brings. Splitting a node inside Aquaculture helps users who find their way there, but it cannot give them a reason to look.

Limitations

This was a self-directed project. I designed, recruited for, ran and analyzed it alone, with no participant panel or budget, to put my research and information architecture skills to the test outside a course. Several limitations below are the cost of that. I have recorded them plainly, because knowing where a study is weak is part of reporting it honestly, and the lessons shape how I would run the next one (see Reflection).

Sample size. Sixteen participants is well below published guidance for quantitative tree testing. Lyssna recommends 40 to 60 participants with 30 as a minimum, and at least 30 per version when comparing structures (Lyssna, n.d.). Maze cites a recommendation of at least 50 (Maze, n.d.), and a Nielsen Norman Group article notes that comparing trees typically needs 50 or more participants per tree (Laubheimer, 2024). At 16, a shift of two or three participants moves the headline rate by more than ten points. The findings are directional, supported by path-level evidence, and should not be read as reliable effect sizes.

Learning effects and uneven gaps. Fifteen participants completed the same task twice, with gaps between sessions ranging from four to 23 days. That range was not planned. After 11 baseline sessions, I judged the sample too small and recruited five more, which stretched the baseline over nearly three weeks. The revised round then ran within a week for everyone, so early participants had about three weeks to forget the task and the last ones about four days. Some will have remembered hunting for ecosystem content, which pushes round-two results upward; this is the most likely explanation for the rise in Aquaculture entry that the revision cannot account for. Extending recruitment was the right call for the sample, but it should have been planned before the first session.

No control task. I excluded my team's second task, which leaves no measure of whether the revision affected findability elsewhere. At the time, a control seemed low-value: the change stayed inside one branch and moved nothing out of it, and with 16 participants a control task could only have detected a large effect. Both points still hold. But the split did lengthen Aquaculture's list from five options to six, which could in principle make its other pages slightly harder to find, and only a task aimed at one of those pages would show it.

Two changes at once. The revision altered both structure and labels simultaneously. The improvement cannot be attributed to one or the other. A study designed for attribution would have tested the split and the relabel separately, at a cost in participants this project could not meet.

The baseline did not replicate the original. My team recorded 6 of 8 participants succeeding on this task; this study recorded 4 of 16. A gap of 50 percentage points is too large to put down to sampling alone, and no single explanation covers it. Likely factors include different participants, different recruitment, different briefs, and small samples on both sides. The comparison within this study is unaffected, because both of my rounds used the same task, brief and procedure. The original figures should be read as historical context, not as a benchmark.

Participant composition. Participants were students, family and personal contacts, not DFO's actual users. Commercial fishers, aquaculture operators, Indigenous fisheries staff and job seekers arrive with vocabulary and expectations none of these participants have. Participants also role-played the task: none was a marine animal safety activist, and someone who is would likely arrive with different words in mind and a firmer sense of what aquaculture and fisheries cover. The same is true of my team's original participants. Nearly half the sample had also previously studied this specific website, which is a form of exposure no real user has.

Brief departure. The brief defined *fisheries* and *aquaculture*, which my team's original brief did not. The reasoning is set out under Methodology. It does not affect the baseline-to-revised comparison, since both rounds used the identical brief, but it does rule out direct comparison with the original study's numbers.

Unmoderated sessions. Participants were not observed and did not think aloud, so the data shows where they clicked but never why. The explanations in this report are inferences from path patterns and label semantics. A moderated follow-up would test whether participants' reasoning matches those inferences.

Data quality. One participant began the baseline three times before completing it, so their counted response follows two prior exposures to the task. One baseline result set collapses two identical paths into one row, leaving those two participants individually unidentifiable. Neither could be corrected after the fact, so both are noted here.

Method boundary. Tree testing removes search, visual hierarchy, homepage promotion and in-page cross-links. A real user facing this task would likely use site search and might never encounter the navigation problem this study measures. The findings describe the navigation structure, not the website as experienced.

Recommendations

Four recommendations follow from the data, ordered by evidence strength.

1. Adopt the split, with further testing. Separating fish-health content from ecosystem content raised success among participants who opened Aquaculture from 40% to 69%, and every participant who reached the environmental node chose correctly. The change is contained within one branch and moves no content out of Aquaculture. It should be re-tested with at least 30 participants before implementation: the direction is consistent, but the size of the effect is not yet established.

2. Resolve the competition between Aquaculture Science and the environmental node. *Aquaculture environmental risk assessments* tied for the most common failure destination after the revision, and it is the only one inside Aquaculture. The most defensible option to test is sharpening the *Aquaculture Science* label so it reads clearly as research, which is what the page is. Moving the page under the environmental node is harder to justify for that same reason. Either change can be tested on its own against the revised tree.

3. Test a cross-link into Fisheries Management. Four baseline failures and two revised failures ended in Fisheries Management's pages. Listing *Protecting Aquatic Ecosystems* in both locations would serve users whose first instinct is the management category, without removing it from Aquaculture. Nielsen Norman Group advises using cross-listing only where data shows a substantial share of users look in more than one place, and warns that overuse "tends to dilute the clarity of an information architecture" (Laubheimer, 2024). Given the taxonomic objection set out under The revision, it should be tested on its own before it is adopted.

4. Treat the absence of environmental content as a content question, not a navigation one. A review of the full site inventory found only two pages with clear ecosystem-protection content, *Protecting Aquatic Ecosystems* and *Ghost Gear*, plus related funding programs under Aboriginal Fisheries. No navigation change can give this topic a coherent home while there is so little content to put in it. That finding belongs with DFO's content team.

Implementation considerations

A real DFO content team would raise objections that this study is not positioned to resolve. They are recorded here because a recommendation that ignores them is incomplete.

URL migration. Splitting one page into two parent nodes affects existing URLs, inbound links and search ranking. Redirects are routine but not free, and the work belongs in the implementation estimate.

Label review. *Environmental Impact and Reporting* was chosen partly to avoid *environmental protection*, which carries specific meaning in Canadian federal legislation. Final wording would need communications and legal review, and the tested label may not survive it.

Terminology consistency. DFO already publishes an *Aquatic animal health science* section elsewhere on its site (Fisheries and Oceans Canada, n.d.-a). Before a *Fish Health Management* node goes live, its name should be checked against that section so the site does not use two different names for related content. This did not need checking before testing: the tree test measured findability within the tested tree, which did not include that section.

Content ownership. Government web pages are typically maintained by a named program team responsible for keeping them accurate. Today one team owns *Environmental Managing and Reporting*. Splitting it creates two parent pages, so DFO would need to decide whether that team keeps both, or whether the fish-health page moves to a team that manages fish health elsewhere. This is an organizational decision more than a design one, but a page without a clear owner tends to go out of date.

Noted but not recommended

Aboriginal Fisheries uses terminology that most Canadian federal content has moved away from in favour of *Indigenous*. This was not tested and is not a findability finding, but it was observed during the inventory and is worth flagging to a content team.

Reflection

The most useful thing I learned is that part of the explanation for my team's failed revision was in our own earlier research.

We ran a card sort and wrote up the finding that no participant grouped aquaculture with enforcement, management, commercial, seal or shellfish content. One assignment later, we renamed the aquaculture category to make its meaning clearer, and I was part of that decision. At the time it was a sensible response to what we saw: participants were choosing Fisheries Management over Aquaculture, so we assumed they did not understand the word. Only after reflecting on both assignments together did a better explanation emerge. We had treated the card sort and the tree test as separate deliverables when they were two halves of one investigation.

That is the habit I would most want to change in how I work. Generative research explains why evaluative research produces the results it does. Reading them together is where the diagnosis comes from.

Failed results are results

My team's revision reduced success on both tasks. At the time that felt like the project going wrong. It was the most informative thing in the dataset: it ruled out the wording explanation by testing it, which is what let me look for a different cause a year later.

I applied the same reasoning to my own study. When the revised round showed higher success but barely improved directness, the temptation was to lead with the number that looked best. The directness figure is the one that says something true about what the fix actually did, and the report leads with both.

Predicting before measuring

Writing down a predicted figure before building the revised tree changed how I read the results. Because I had committed to *which* failures the fix should recover, I could check the mechanism as well as the outcome. A rising success rate shows that something worked. A rate that rises for the reasons I predicted suggests I understood why.

It also made it harder to move the goalposts. Missing the directness threshold is recorded in this report because I set it in advance and cannot now argue it did not matter.

What I would do differently

Recruit properly, and before starting. Sixteen participants is likely the weakest part of this study, and I knew it while running it. I also recruited in two waves, which is what made the gaps between rounds so uneven. A target of 30, recruited before the first session, would have made the comparison more than directional and kept the gaps consistent.

Keep a control task anyway. I dropped the second task because the change was contained in one branch and the sample was too small for a control to catch anything but a large effect. That was a defensible call, but it left a question I can only answer with reasoning, with no data behind it. One extra task would have cost about thirty seconds per participant.

Separate the variables. I changed structure and labels together, so I cannot say which did the work. With more participants I would have tested them independently.

Space the rounds further apart. Gaps of four days are too short to be confident that participants had forgotten the task.

On restraint

The decision I was least certain about, and most glad I made, was not fixing everything. There were three other changes I could have justified from the data. Making them all would have produced a tree I liked better and a result I could not explain.

The narrower change produced a smaller improvement and a clearer answer. For a study whose purpose was to test a diagnosis, I think that was the right trade.

Sources

- Fisheries and Oceans Canada. (n.d.-a). *Aquatic animal health science*. https://www.dfo-mpo.gc.ca/science/aah-saa/index-eng.htm
- Fisheries and Oceans Canada. (n.d.-b). *Environmental management and reporting*. https://www.dfo-mpo.gc.ca/aquaculture/environment-environnement-eng.html
- Laubheimer, P. (2024). *Tree testing part 2: Interpreting the results*. Nielsen Norman Group. https://www.nngroup.com/articles/interpreting-tree-test-results/
- Lyssna. (n.d.). *Tree testing guide*. https://www.lyssna.com/guides/tree-testing-guide/
- Maze. (n.d.). *How to run a successful tree test*. https://maze.co/guides/tree-testing/process/
- Sauro, J. (2013). *Using tree-testing to test information architecture*. MeasuringU. https://measuringu.com/tree-testing-ia/

Appendices

A. Task wording

Identical in both rounds:

> "You are a marine animal safety activist, and you are looking to find information on minimizing damage to undersea ecosystems. Where would you find this?"

Correct destination: *Protecting Aquatic Ecosystems*.

B. Participant brief

Identical in both rounds:

> This website is run by Fisheries and Oceans Canada, the federal department responsible for Canada's fisheries, oceans, and waterways. People use the site to find information about fishing licences, regulations, the fishing industry, conservation programs, and careers in the field.
>
> You'll see a menu of categories that you can click to explore the options underneath each one.
>
> I'll give you a short scenario and ask where you would go to find the information you're looking for. Navigate through the menu as you normally would. If a path doesn't seem right, go back and try another one.
>
> There are no wrong answers. I'm testing how the website organizes its information, not you, so please follow your first instincts.
>
> Please note: The tree test is timed, so make sure you're focused and ready before you begin.
>
> **Definitions**
>
> Fisheries: The catching, harvesting, or farming of fish and shellfish within a particular area or industry.
>
> Aquaculture: The farming of fish, shellfish, and aquatic plants in controlled environments.

C. Baseline tree structure

The tested tree is a 60-node subset of the 115-node site inventory, selected by my team as the nodes relevant to the tasks being tested. Pruning a tree this way is standard practice for keeping a tree test focused. Labels appear exactly as tested.

- **Aboriginal Fisheries**
  - Food, social and ceremonial fisheries
  - Fishing in pursuit of a moderate livelihood
    - Fisheries science and advice
- **Commercial Fisheries**
  - Commercial fishing licences

    - Exporting and importing fish
    - Fisheries management
    - Commercial fishery requirements
- **Recreational Fishing**
    - Provincial and territorial fisheries
    - Maritimes region recreational fisheries
- **Fisheries Management**
    - Fisheries enforcement
        - Enforcement news
        - Fisheries act convictions
        - High-seas monitoring
        - Fishery officer jobs
        - Ghost gear
        - Fishing regulations and requirements
    - International management
        - Monitoring, control and surveillance
        - Governance and diplomacy
        - Canada and international fisheries
    - Policies and frameworks
    - Fisheries management decisions
        - Integrated fisheries management plans
        - Fisheries openings and closures
- **Seals and Sealing**
    - Identify a species
        - Ringed Seal
    - Ensuring the seal harvest is humane
- **Shellfish Harvesting and Safety**
    - Shellfish harvesting map
    - Reasons for closures
    - Contact our offices
- **Aquaculture**
    - Aquaculture in British Columbia

        - Public reporting
        - Aquaculture licensing
    - Business and Licences
        - Licensing in most provinces and territories
        - Aquaculture sites and maps
        - Licences for introductions and transfers
        - Communities and employment
    - Environmental Managing and Reporting
        - Reducing Disease Risks
        - Managing Diseases and Parasites
        - Protecting Aquatic Ecosystems
        - Aquaculture Public Reporting
        - Treating Outbreaks
    - Governance, Regulations, and Reporting
        - Roles and Responsibilities
        - Aquaculture Collaboration
        - Aquaculture Activities Regulations
    - Aquaculture Science
        - Aquaculture regulatory research
        - Aquaculture collaborative research and development
        - Closed containment
        - Aquaculture environmental risk assessments

D. Revised tree structure

Identical to the baseline in every branch except Aquaculture, where *Environmental Managing and Reporting* is replaced by two nodes.

- **Aquaculture**
    - Aquaculture in British Columbia *(unchanged)*
    - Business and Licences *(unchanged)*
    - **Fish Health Management** *(new)*
        - Reducing Disease Risks
        - Managing Diseases and Parasites

        - Treating Outbreaks
    - **Environmental Impact and Reporting** *(new)*
        - Protecting Aquatic Ecosystems
        - Aquaculture Public Reporting
    - Governance, Regulations, and Reporting *(unchanged)*
    - Aquaculture Science *(unchanged)*

E. Raw data

Full path exports for both rounds, including each participant's click sequence and completion time, were exported from ProvenByUsers and are available on request. The failure destinations and funnel figures in this report are derived from those exports.