CASE FILE 001 · OPEN INNOVATION · ENTERPRISE AI · 2026

A design system that remembers why

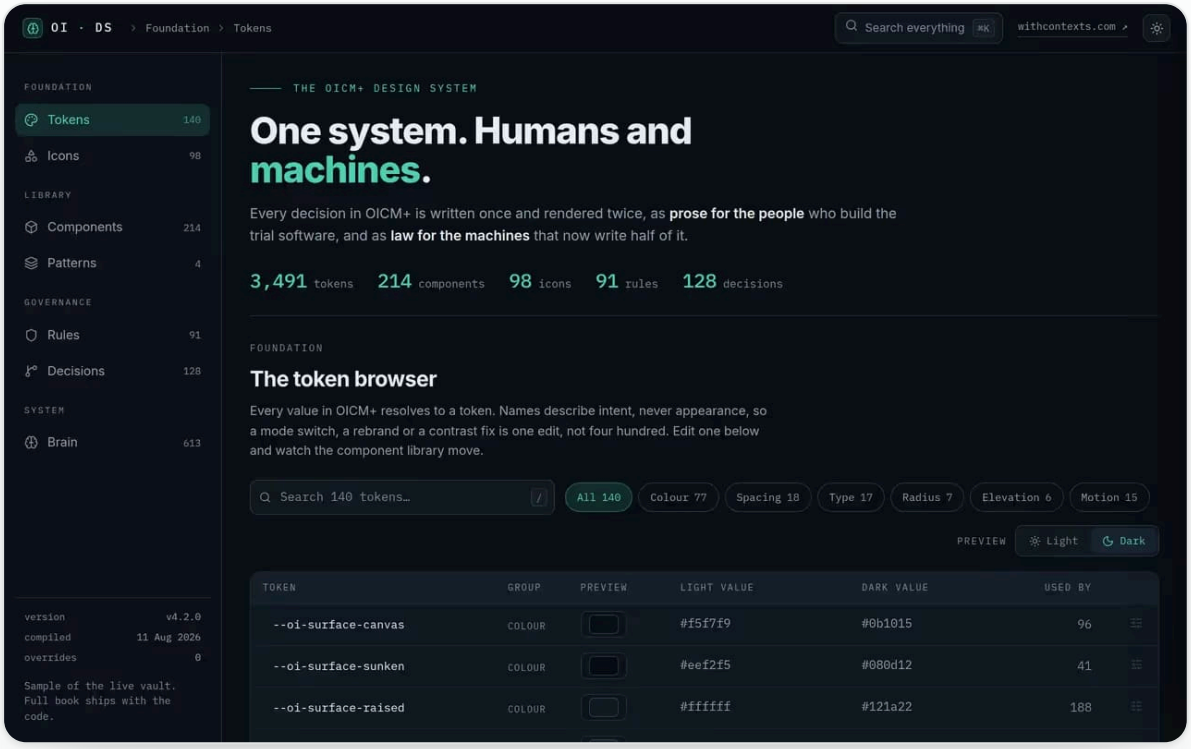
WHAT	the design system and AI prototyping pipeline of an enterprise ML platform
FOR	five product teams and nine designers
PROBLEM	the product ran on unconfigured Ant Design, with four heights for one bar chart

A token-governed component system, an **institutional memory** read by humans and AI agents alike, and a pipeline that turns plain English into working prototypes. Built as product design lead, hands-on across AI, agents and workflows, with the frontend team.

OI·DS · the design system portal

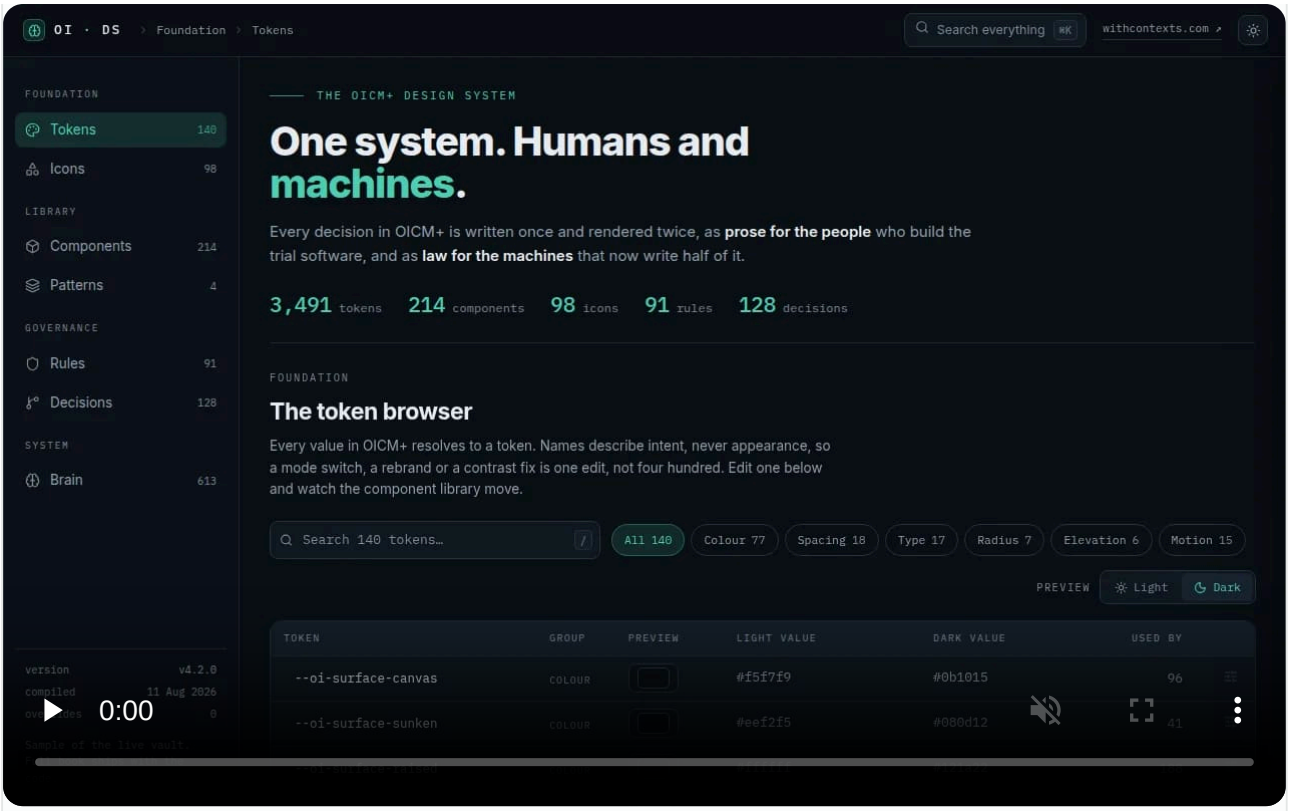
Accessibility: The portal passes axe-core's WCAG 2.1 AA rules at desk and phone, in both settings, and in the 3 states this file links into; the Storybook below it is the repository's own build and is not graded here. Not yet tried with a screen reader. The statement

Weight on a first visit: the portal, 60 KB compressed, against a budget of 65 KB. The build fails if a product passes its budget. It is one page, its script and style inside it.



The portal is built for a desk. Open it full screen to use it properly.

Open OI·DS ↗

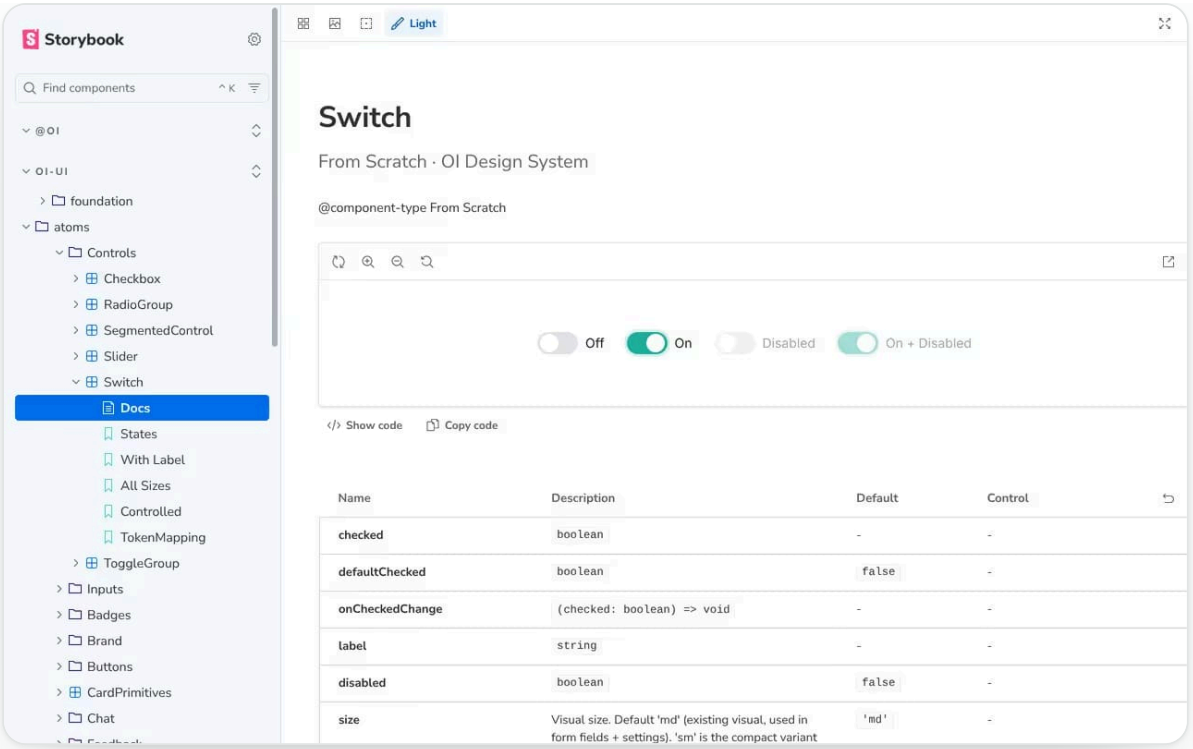


A walkthrough, 21 seconds, recorded from the rebuild. The portal, tab by tab: the tokens, a theme switched, the components, the rule book in both its forms, and the decisions log. Each step is captioned as it happens, and the transcript has the same words.

Watch while you read

TRANSCRIPT

OI·DS · the Storybook



The Storybook is built for a desk. Open it full screen to use it properly.

Open the Storybook ↗

ROLE	COMPANY	TIMEFRAME
Product Design Lead	Open Innovation (enterprise ML platform)	2026 · 100+ shipped releases

I OWNED	The design system end to end (tokens, components, documentation and governance), the DS-Brain memory, and the AI prototyping pipeline
WITH	The Head of Frontend Engineering and the frontend team, for five product teams and nine designers
THEIRS	The products built on the system, by five product teams and nine designers, and the frontend code the Head of Frontend Engineering's team writes against it
CONSTRAINTS	Ant Design stayed underneath and could not be ripped out, and five product teams kept shipping through 100+ releases, so the system replaced its surface token by token, never in a cliff-edge release
FRICTION	Getting engineers to want the system, which was harder than building it; the rule book and the memory reached them after the library, and should have come first

THE OUTCOME

Nobody waits for a designer

Plain English in, a working prototype in the production components out, in about 15 seconds.

THE PROBLEM

A platform wearing raw defaults

An enterprise ML product running on unconfigured Ant Design: four heights for one bar chart, !important ×70.

THE MOVE

Tokens, rules and a memory

3,396 tokens in a governed chain, 91 numbered rules, and a 378-document brain read by people and AI alike.

FIG. 001 · THE COMPONENT LIBRARY

231 POINTS · ONE PER COMPONENT

231 COMPONENTS

| AS OF JUL 2026 Every one resolves through the token chain. Nothing in the library owns a colour

JOINED TO ITS NEIGHBOURS

A change to one is a change to everything touching it, which is the problem the rules exist for

01

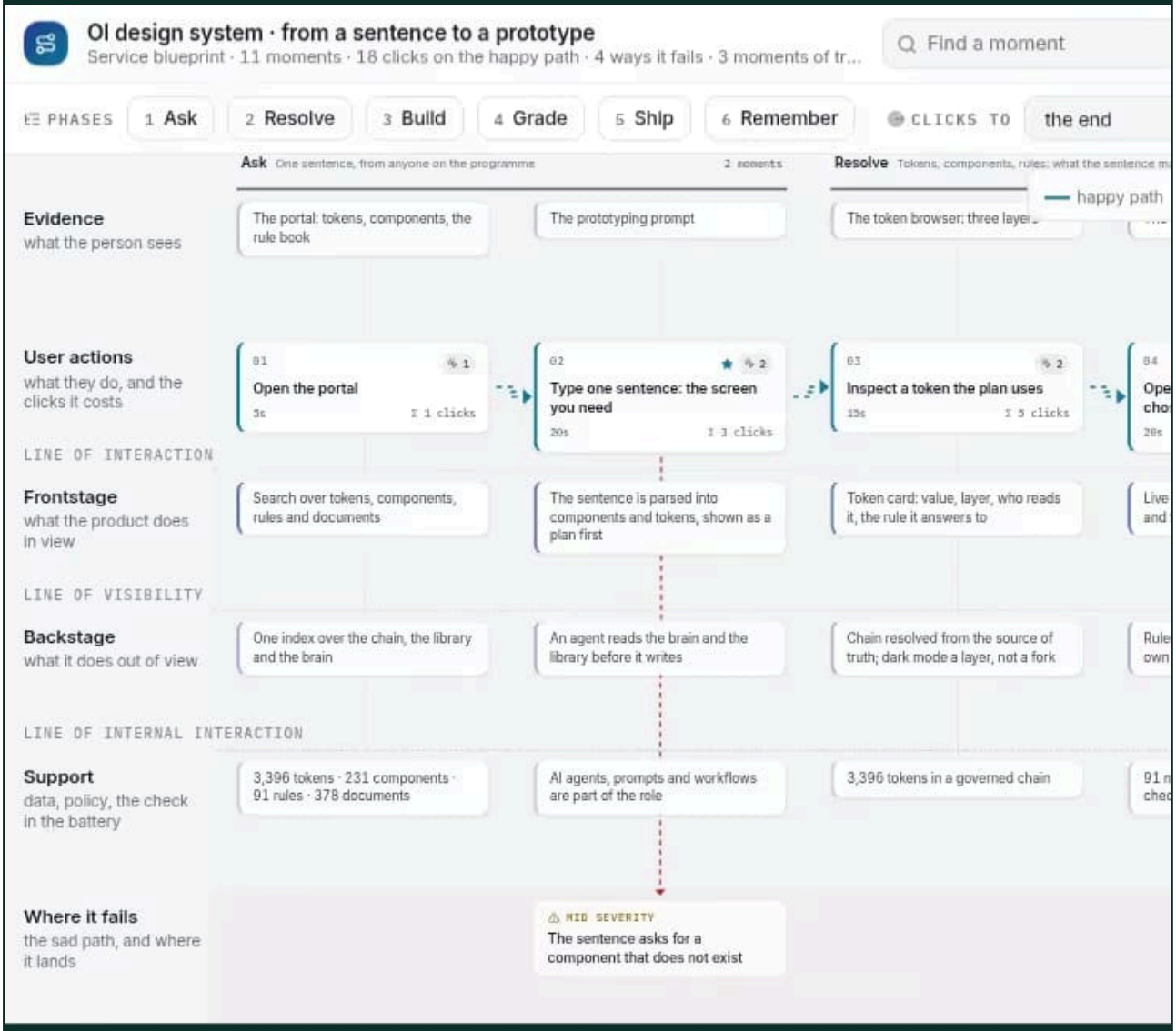
THE BEFORE AND AFTER

The same dashboard, two years of debt apart

We didn't rip Ant Design out, the system replaced its surface token by token, wrapping it in OI-prefixed components, so users never met a cliff-edge release. Drag the line.

SERVICE BLUEPRINT

From a sentence to a prototype as a service: eleven moments from a sentence to a graded, shipped prototype, and the brain that remembers why. Pan, zoom, pick a moment, count the clicks to any point, switch the sad paths on.



Wheel to pan, pinch to zoom, click a moment for its detail. ← → walk the moments, f fits the map.

Open the blueprint full screen →

FOUNDATIONS · 3,396
TOKENS · 6 LAYERS

02

No hex above this line

Every visual value resolves through a three-layer chain, and a component may never touch a raw value. That one prohibition is the whole

trick: | AS OF JUL 2026

PRIMITIVE
The raw value
a colour, a size, defined exactly once

→ SEMANTIC
What it means
"accent", "canvas", "danger", intent, not ink

→ COMPONENT
What a button reads
components see only this layer, 3,396 variables deep

→ THEME SWAP
Flip the primitives
every component follows, dark mode cost zero component code

Above the tokens sits the six-layer atomic hierarchy, and skipping a layer **fails the build**, not the review. The rest of the foundation is deliberately small: 24/12 brick spacing, one chrome accent, Inter for UI and Plex Mono for data, two motion curves.

The chain, as the source has it

Followed from name to name until each one reaches a value, the 3,396 tokens in the stylesheets fall like this. | AS OF JUL 2026

741 hold a value themselves: the primitives, and the sizes and weights that need no name above them

1,403 point at one name that holds a value

1,008 are two steps from a value

233 are three or more steps from a value

11 reach a name no stylesheet in this repository declares

ONE CHAIN, END TO END

`--oi-fd-color-text-primary` → `--oi-color-primary-100` → `--oi-teal-300` → `#84ebdd`

FIG. 002 · THE TOKEN CHAIN



Components reference roles, never values, which is the whole reason a theme swap costs nothing below it.

A system that remembers why

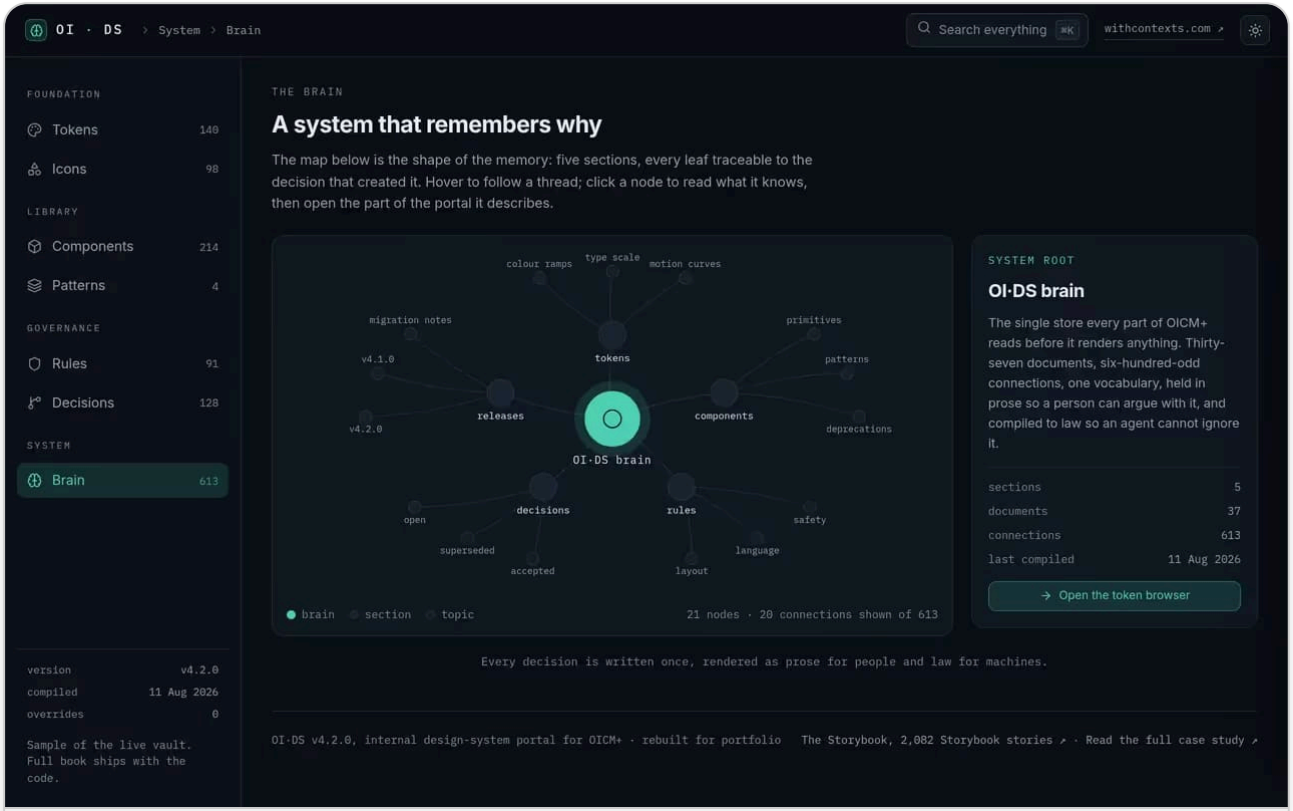
Institutional memory, treated as a product: written once, rendered twice, prose for people, structured rules for machines, so a new engineer and an AI agent read the same canon before touching a component.

378 docs · 6,683 links

Its spine is an append-only decisions log. Nothing is edited or deleted, a reversed decision is a new entry pointing at the one it overturns, reasoning attached. Ask "why is it like this" and you get an answer with a date on it.

COMPUTED FROM THE VAULT MANIFEST · 17 SECTIONS

The vault itself is the one artifact this site does not serve: it holds the team roster, the full decisions log and the roadmap, and its own security note bars it from public deploys. That refusal is part of the design.



The brain, running: documents and decisions, read by people and agents.

1 · ASKED	2 · THE CANON	3 · THE LOG	4 · ANSWERED	5 · CHANGED
Why is it like this?	**Read the same rules**	**Find the decision**	**With a date on it**	**Only by a new entry**
By an engineer, a designer or an agent, before a component is touched.	Written once and rendered twice: prose for people, structured rules for machines.	Append-only: nothing is edited or deleted, and a reversal points at what it overturns.	The rule's number, the decision behind it and when it was taken.	A rule is followed or retired in the log, and a retired number stays vacant.

04

GOVERNANCE

91 rules, numbered like case law

FIG. 003 · THE RULE BOOK

91 NUMBERS · 89 LIVE

RULE 4

No hex above this line. One prohibition, and the whole chain follows from it

TWO OPEN SQUARES

Retired rules. The numbers stay vacant, so the sequence carries its own history

Two rules were retired and their numbers stay vacant, the sequence carries its own history. Change ships as numbered waves, W11 past W110; there has never been a big-bang release.

- ✓ **The eight-file contract**
  types, tokens, styles, stories and the rest, miss one file, no promotion
- ✓ **Screenshot chrome checks**
  every visual PR compared against the last, not remembered
- ✓ **A fresh-context AI verifier**
  grades finished work cold, citing rule numbers in its findings

Who is allowed to decide what

A rule book says what good looks like. It does not say who gets to change it, and that second question is the one that decides whether a design system survives its author. So the governance file carries a decision-authority matrix: eighteen classes of change, each resolved to the one role whose sign-off the pull request cannot merge without.

One approver, never two. The first operating principle in that file is blunt about why: **multiple approvers diffuse ownership and are non-conformant**. An optional reviewer may be asked and cannot block. The shape is the argument. A component owner adds an atom without asking anyone; nobody but the DS owner can add an iron rule, and that row has no optional review at all, because a rule anyone can add is not a rule.

Four decisions, from the log

Rule three of the same file: any decision that locks a pattern, picks between alternatives or changes a rule produces an entry in the log, and **reverts are recorded**. That last clause is the one that costs something. These four are in the permanent record under my name.

`W17 · 2026-05-19`

Six named roles, and a matrix that resolves change to approver

The brain was a single contributor's notebook. Handing it to a twelve-person team meant defining and locking six roles, writing the authority matrix above, and separating the no-compile rule, which governs ad-hoc work, from CI, which does not get to skip anything.

`WHAT IT COST`

A personal-voice scrub across the whole corpus. The writing stopped being mine, which is the point and still the part I liked least.

`W19 · 2026-05-21 · SUPERSEDES W11 AND W17`

Bricks frame themselves; the canvas never pads

Page-level padding came off the canvas and into every brick, as a per-component token. The paddingInline prop came off a layout component's public API entirely.

`WHY`

Per-template opt-outs had left a hidden contract: a brick that looked right in one page template looked broken in another, and story-level migration broke every time. The fix retired two of my own earlier decisions and changed a public API, which is the expensive kind of correct.

`W17-REVERSAL-CANCELLED · 2026-05-21`

A system-wide mandate, introduced and reverted the same session

Earlier that day I locked a canvas-gutter mandate across the system, hardened a rule to enforce it and added a guard grep. Hours later I reverted all three.

`WHY`

The mandate was based on a wrong reading of the architecture. Bricks carry their own padding precisely because they are the frame; a gutter on top of that compresses the content. The entry exists so that the next person to have the same idea finds out it was tried.

W110 · 2026-06-30

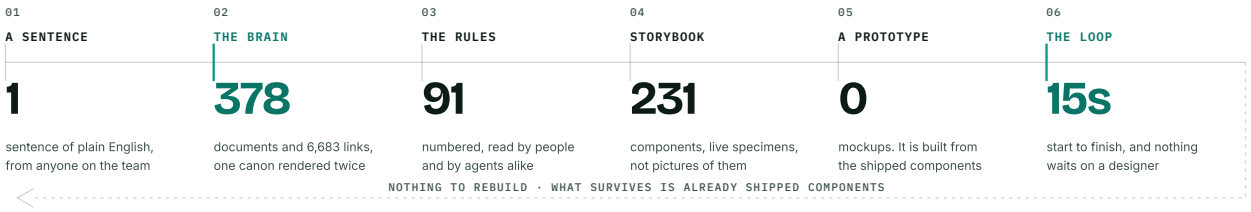
Killed the prototyping route we were already building

Five options were compared in a written strategy memo. The one already under construction was killed because its sandbox could not import the real component library, so every prototype it produced was a CSS imitation of the system rather than the system. The feedback that settled it was not diplomatic: not our components, not our style.

WHAT THE DECISION DID NOT DECIDE

The model provider, export-to-TSX and authentication, each deferred in writing with a reason. A decision record that does not state its own boundaries is a decision someone else has to make again by accident.

FIG. 004 · THE PROTOTYPING LOOP



The designer left the critical path. Anything that survives the loop is already shipped components.

05

THE AI CHAPTER

We killed the tool we were already building

Five routes compared in a written memo. We killed the one we were already building the day the evidence landed, and the replacement shipped in two phases.

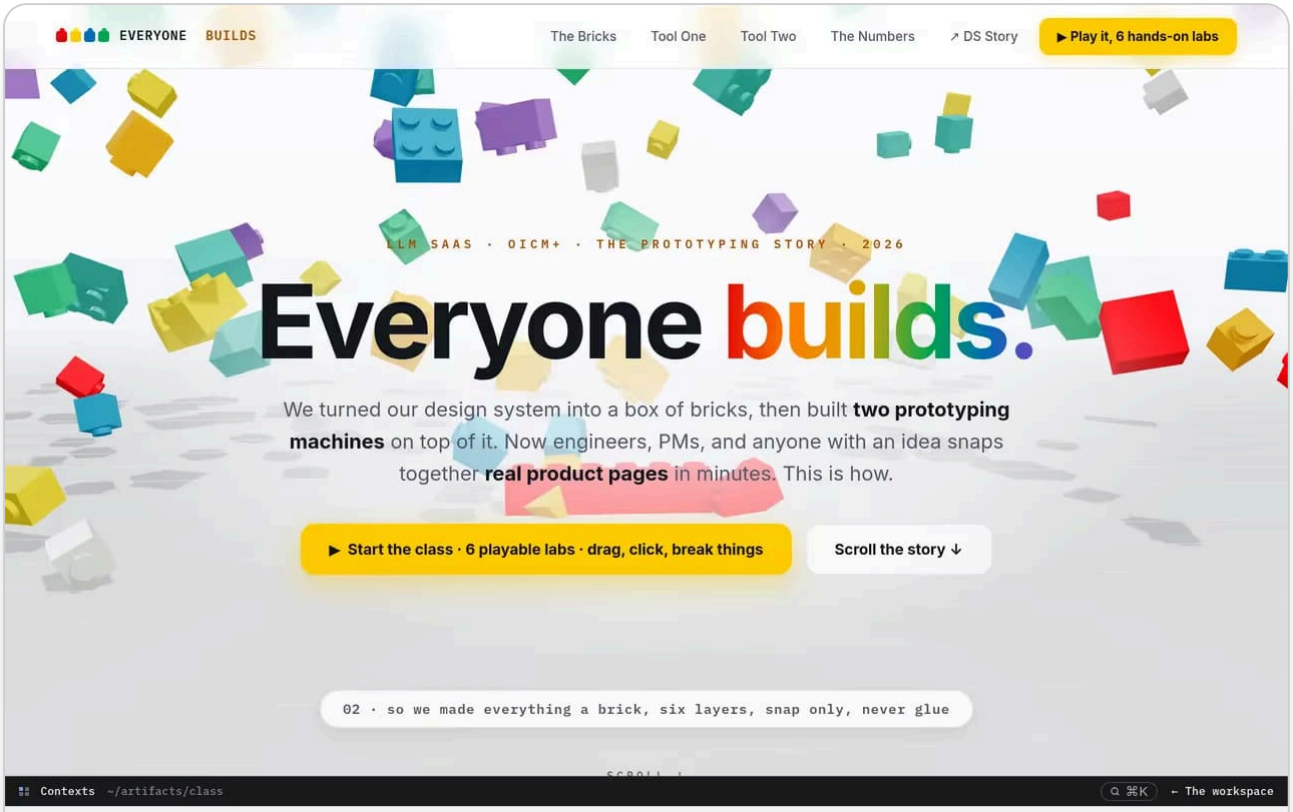
DECISION W110, THE FULL ENTRY, REASONING AND ALL, IS IN THE LOG

A browser sandbox can't import the production component library, so every prototype degraded into a CSS imitation of the system. The post-

mortem went into the log like any other entry, and the replacement renders from the library itself:

PROMPT	→ PIPELINE	→ OUT	→ LOOP
Plain English	**Production components**	**A working Storybook story**	**~15 s per tweak**
a designer or PM types a sentence	rendered from the library itself, not a copy of it	inside the product's own app shell	two dozen production-track prototypes shipped so far

None of this hands taste to a model. The AI builds under the same constraints as any engineer, the rules, the tiers, the cold-eyed verifier, and when it refuses a request because a component doesn't exist, that refusal routes back to design as a **demand signal**. The last piece was teaching: an interactive atomic-design class for the whole company, public below as Everyone Builds.



From a sentence to a prototype: the story, as it was taught to the company.

What kind of AI system it is

Not an agent left to roam: a short, fixed workflow with two checks after it, which is where Anthropic's guidance on agents says to start.

THE PIPELINE, IN THE PATTERNS OF “BUILDING EFFECTIVE AGENTS”

PATTERN	USED HERE	WHERE
Prompt chaining	A sentence goes in, the output is rendered from the production components, and what comes out is a Storybook story in the product’s shell	The prompt, the pipeline, the story
Routing	To a person: when a request needs a component that does not exist, it is refused and the refusal goes to design as a demand signal	The refusal
Evaluator and optimizer	A fresh-context verifier grades finished work cold and cites rule numbers in what it finds; screenshots compare every visual change with the last	The verifier, the screenshot check

How it fails, and what happens next

IT CANNOT REACH THE REAL LIBRARY

The first route ran in a browser sandbox that could not import the production components, so every prototype degraded into a CSS imitation of the system. That route was killed and the post-mortem went into the log.

IT PRODUCES SOMETHING THAT BREAKS A RULE

The verifier reads the finished work without the conversation that made it and names the rule numbers it breaks; the screenshot check catches a visual change nobody asked for.

IT IS ASKED FOR SOMETHING THAT DOES NOT EXIST

It refuses rather than inventing a component, and the refusal is routed to design as evidence that the system is missing a part.

One prompt, read line by line

From the pipeline’s prompt library, as written; the company and the deployments in it are invented. What each line is for:

```
/prototype A deployments list page.
Breadcrumb: Workspaces > Acme Corp >
Deployments. Title "Deployments" with
count badge "108 total". Subtitle
"Active model deployments across all
clusters".

3 KPIs: Running 84 (up 6), Pending 12
(down 2), Failed 4 (down 1).

Then a 6-row table titled "All
Deployments" with columns Name, Status
(badge), GPUs, Replicas, Owner. Use
names like llama-3-70b-prod, embedding-
svc, reranker-stg, translation-en-fr,
speech-to-text-test, image-gen-canary.
Mix statuses.
```

1 **A page type, not a picture**
"A deployments list page" names a kind of page the library has templates for, so the model fills slots rather than drawing a layout.

2 **Where it lives**
The breadcrumb puts the page inside the product's own navigation, so the story renders in the real app shell.

3 **Named parts, in order**
Header, then KPIs, then a table: each maps to a section type in the library, and the order is the page.

4 **Real-looking data**
Plausible names and statuses make the prototype something a person can react to, not a page of placeholders.

5 **What it does not say**
No colour, no spacing, no font. Those belong to the tokens, and a prompt that set them would be asking the model to break Rule 4.

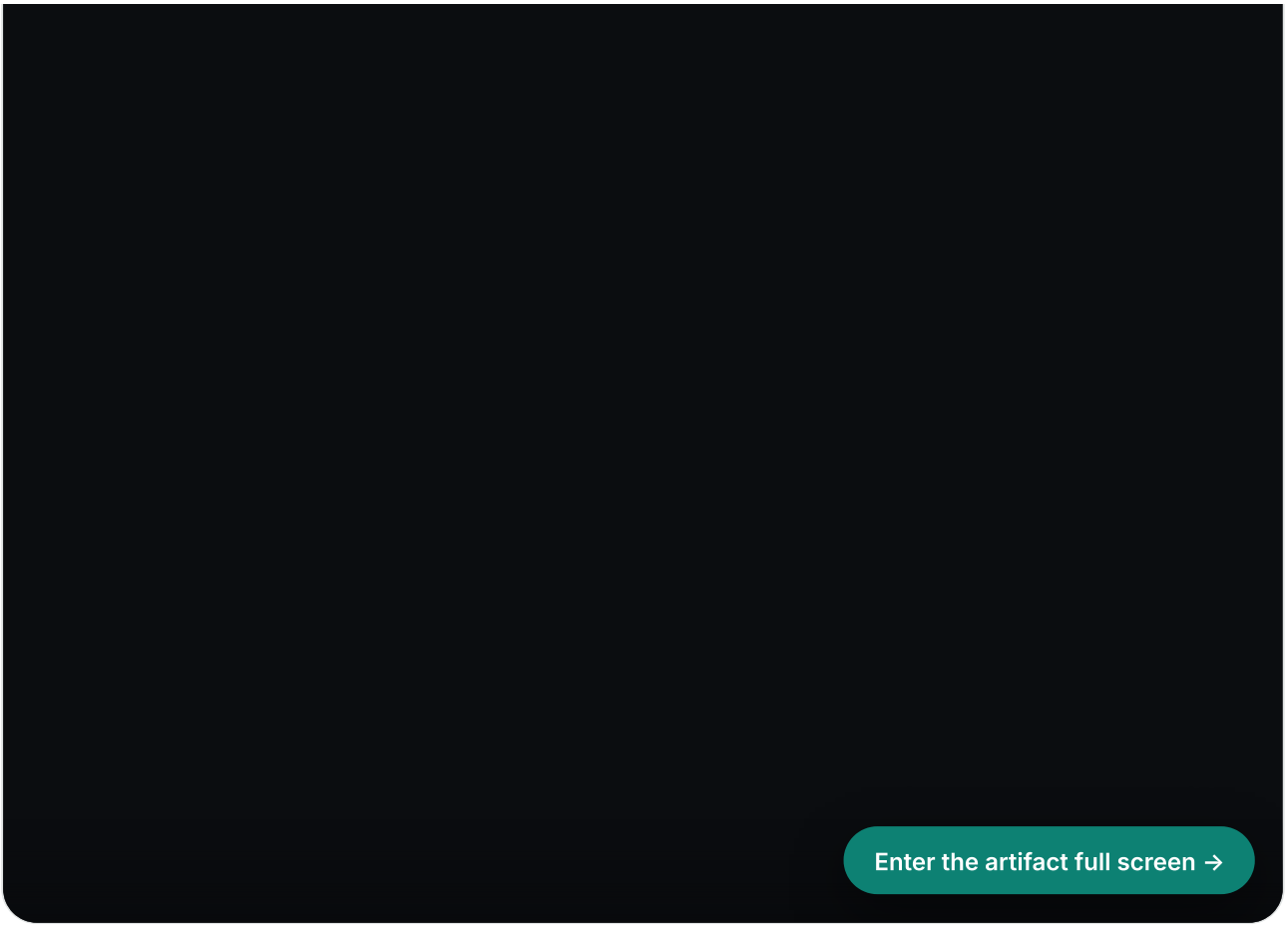
06

THE EVIDENCE

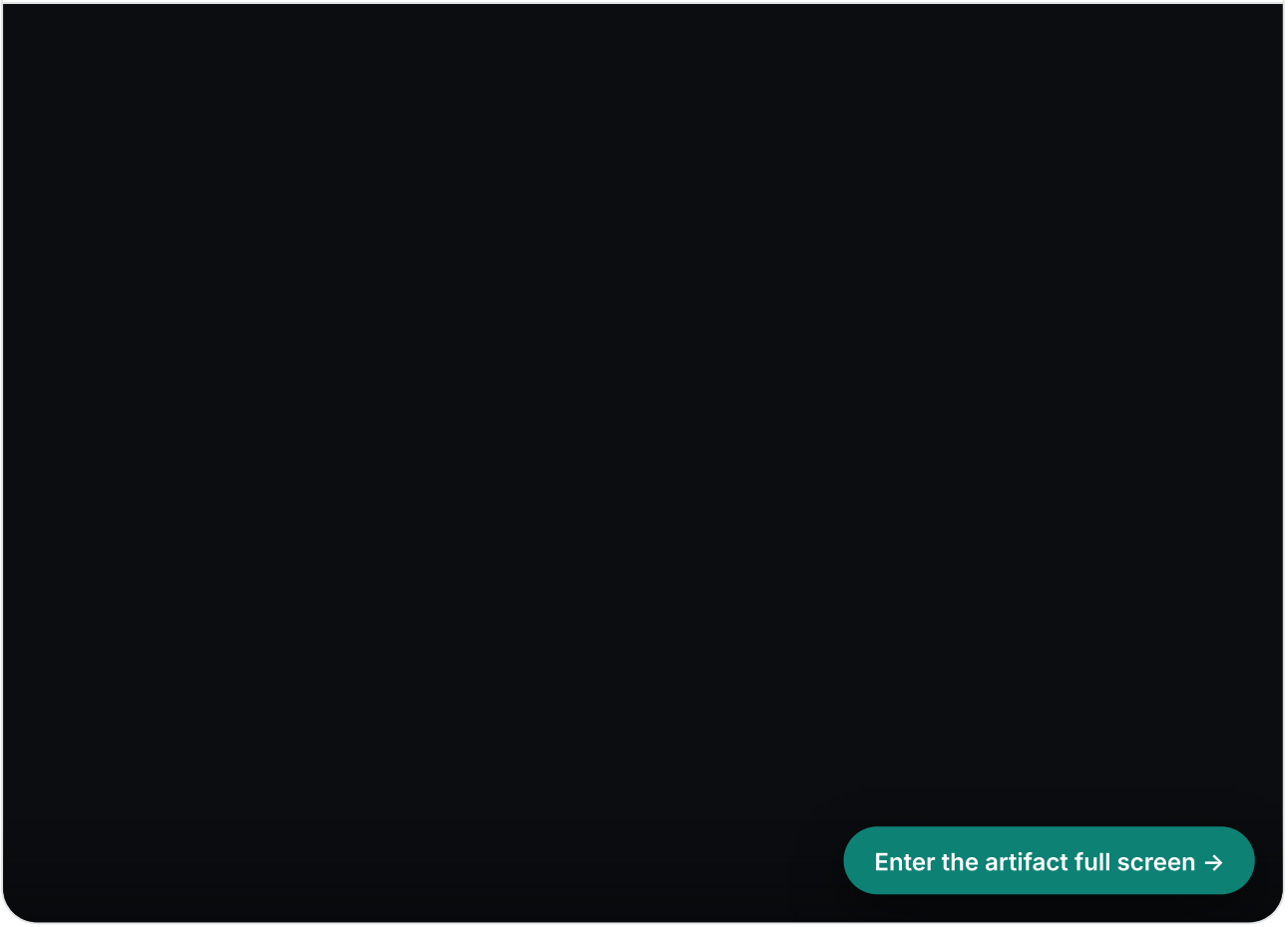
Nothing below
is a screenshot

These are the actual artifacts, the story site in its own tokens, the deck of UX problems that produced the rules, the class that taught the company to build, served from this domain, running live inside this page.

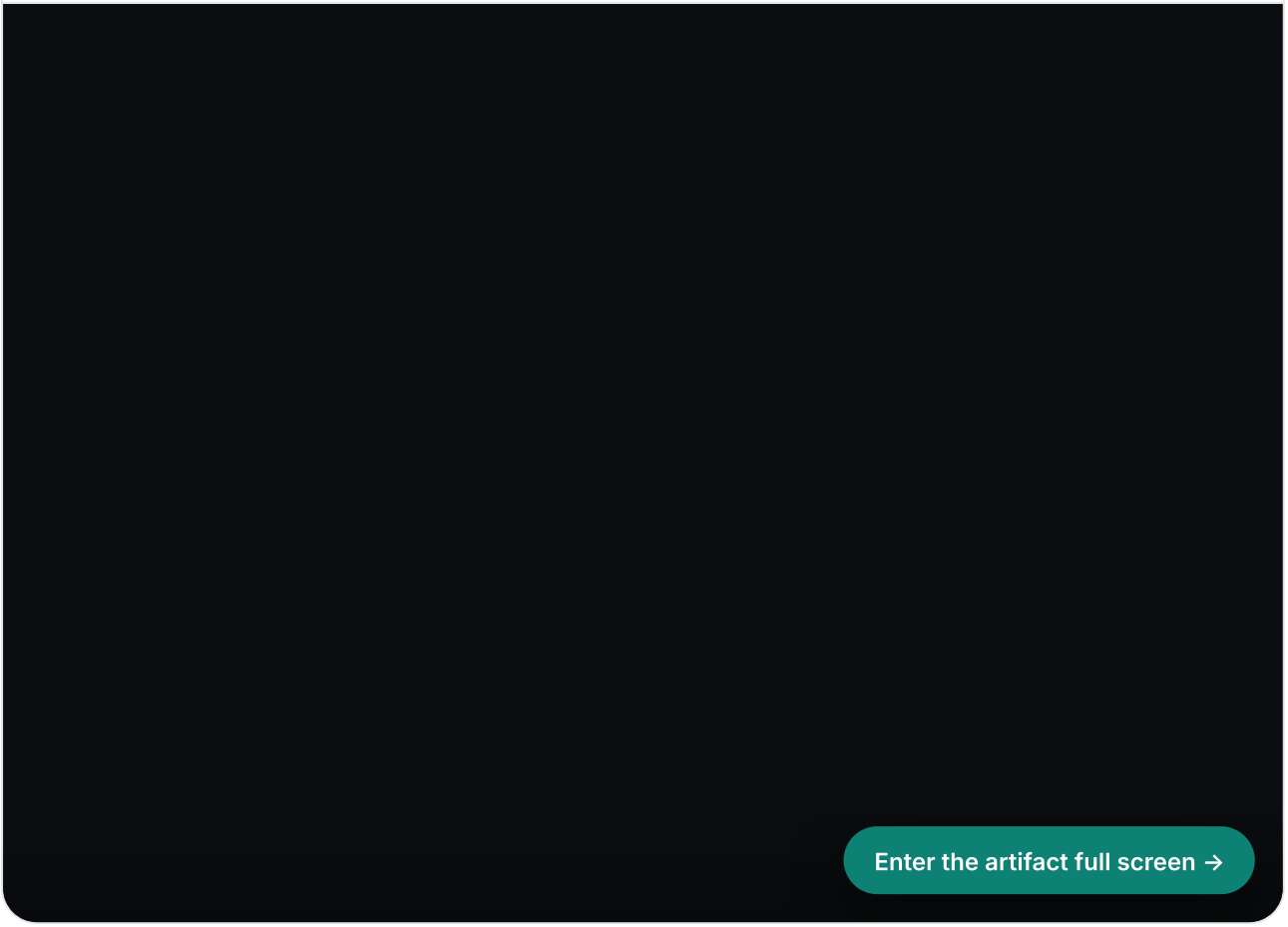
LIVE ARTIFACT — From Chaos to Canon, the story site, in its own tokens. This is the before/after page; drag the line inside it.



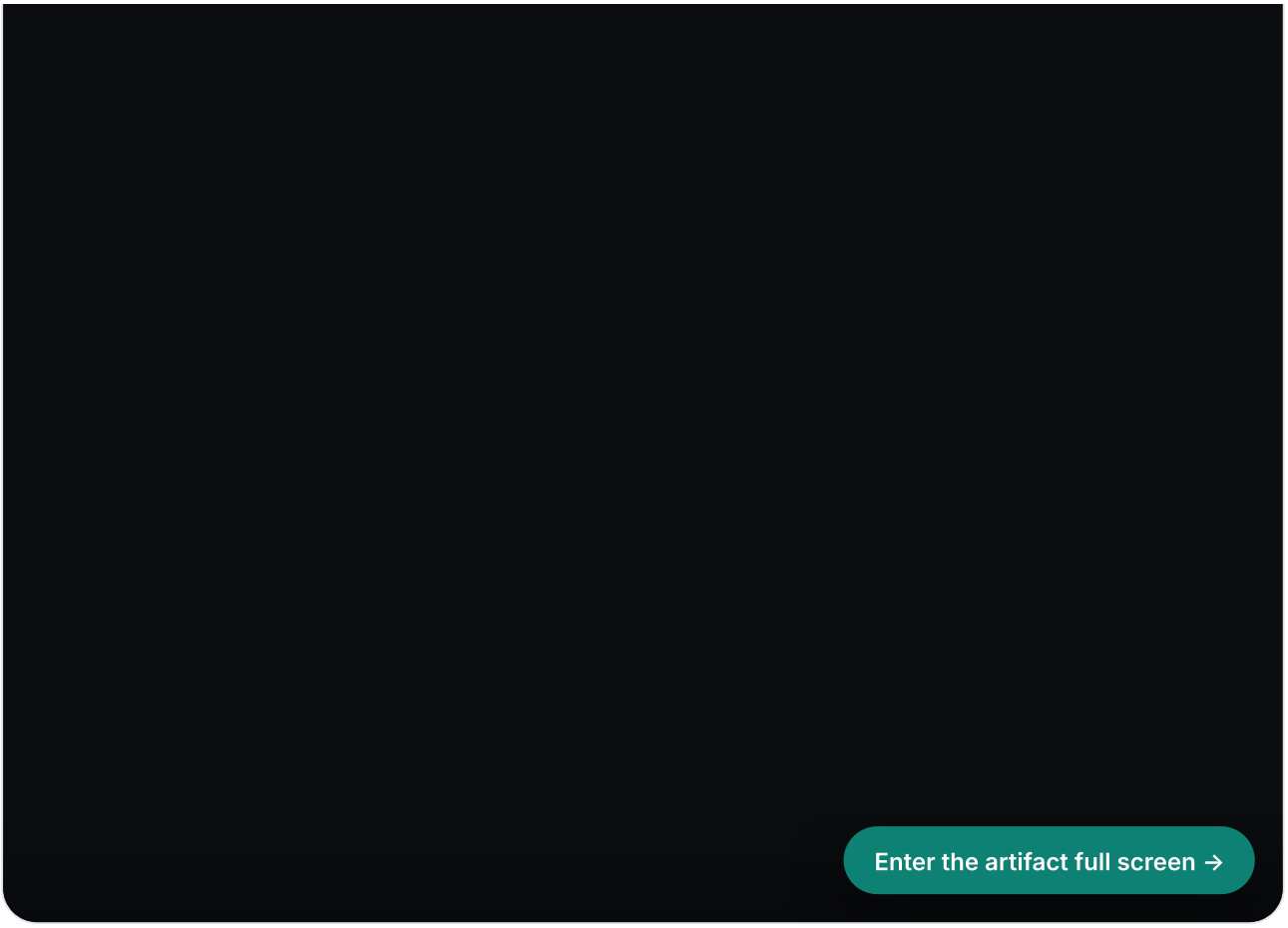
LIVE ARTIFACT The token foundation, 3,396 properties, click any swatch to copy it.



LIVE ARTIFACT The UX Files, 44 slides, 20 playable problems. Arrow keys work.



LIVE ARTIFACT Everyone Builds, the interactive class, six playable labs.



The full story site → | The presentation → | The full story · 8 chapters →

The component library → | The atomic system → | The prototyping pipeline →

The Storybook, open on a docs page →

07

A sentence goes in. Production components come out.

The whole path from a plain-English request to a shipped feature, including the branch we built, measured and killed.

FIG. 005 · THE PROTOTYPING PIPELINE



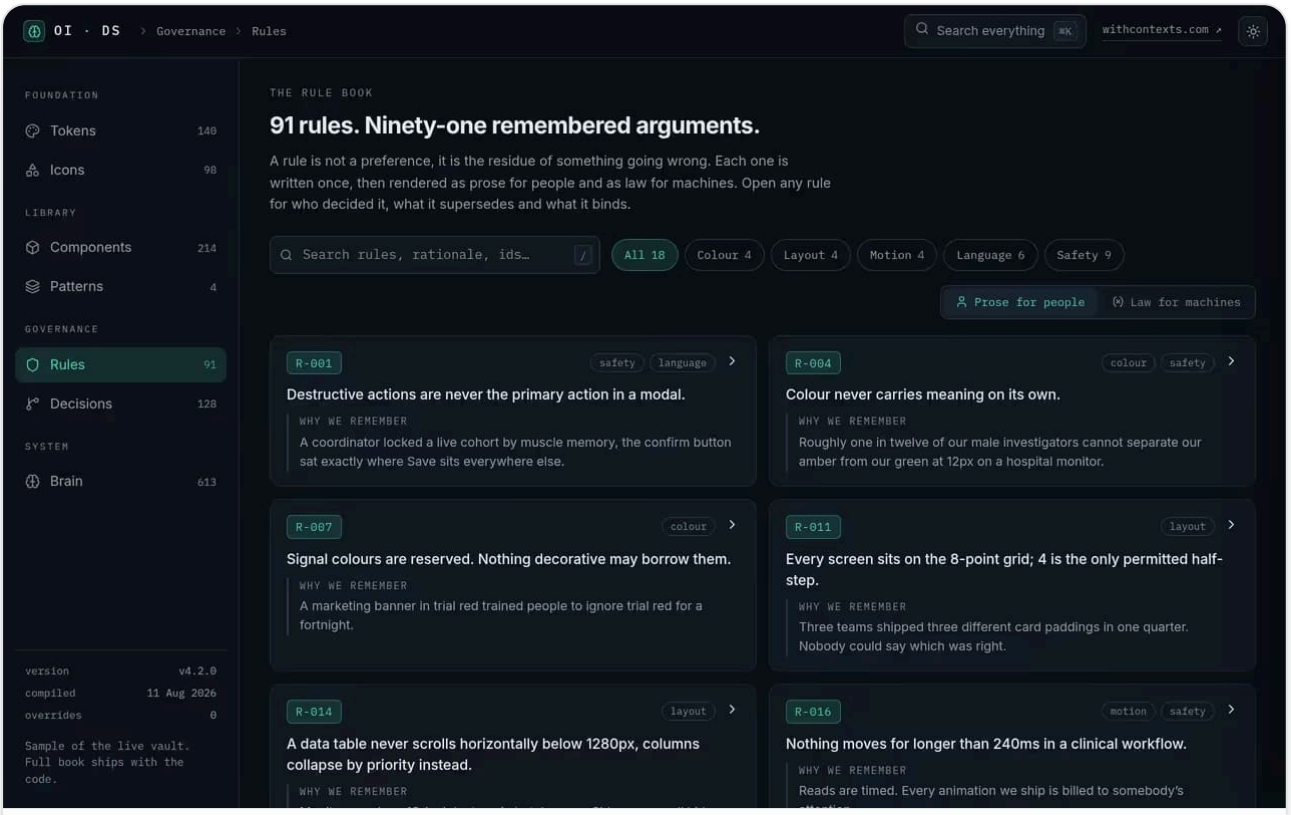
Two branches matter: the verifier’s fail loop, and the dead end at W110 where we killed our own tool.

WHAT WAS BROKEN ·
WHAT CHANGED

08

The debt, itemised, and what each fix moved

Seven things that were wrong when I arrived, the mechanism that fixed each one, and what it changed afterwards.



What moved, in the product: the 91 rules, machine-checked.

01

WAS BROKEN

The platform ran on unconfigured Ant Design. One bar chart carried four control heights, one button family carried three teals, and a single Select wrapper carried **!important seventy times over**.

WHAT I CHANGED

A three-layer token chain, primitive, semantic, component, with one prohibition: a component may never read a raw value. Rule 4, no hex above the token line, fails the build rather than the review.

WHAT IT MOVED

3,396

CSS custom properties now carry every visual value, and dark mode shipped at the cost of zero component code.

COUNTED FROM SOURCE · SEE THE COLOPHON

02

WAS BROKEN

Anyone wanting to try an idea queued behind a designer, and browser-sandbox prototypes degraded into CSS imitations of the system because they could not import the production library.

WHAT I CHANGED

A plain-English pipeline that renders from the component library itself, inside the product's own app shell, after we shipped two phases of our own playground and killed it in a day when the evidence said no.

WHAT IT MOVED

~15 s

per iteration, and roughly two dozen production-track prototypes authored by people who do not write components.

DECISION W110 · THE FULL ENTRY IS IN THE LOG

03

WAS BROKEN

Nobody could say why a decision had been made, so settled arguments reopened every few months and reversed rules crept back in unnoticed.

WHAT I CHANGED

An append-only decisions log at the centre of the vault. Nothing is edited or deleted; a reversal is a new entry pointing at the one it overturns, reasoning attached, dated.

WHAT IT MOVED

378 / 6,683

documents and cross-links, written once and rendered twice, prose for people, structured rules for agents.

COMPUTED FROM THE VAULT MANIFEST

04

WAS BROKEN

Design review was taste against taste, and the finished work was graded by the person who built it.

WHAT I CHANGED

91 rules numbered like case law, cited by number in review, plus a fresh-context AI verifier that grades work cold and quotes rule numbers in its findings.

WHAT IT MOVED

The argument moved from pixels to rules: a disagreement now cites a number, looks up why it exists, and either follows it or retires it in the log. Two rules have been retired; their numbers stay vacant.

FROM THE RULE BOOK

05

WAS BROKEN

One table job was being done by dozens of near-identical implementations, ten filter patterns, eight pagination types, nine footer variants.

WHAT I CHANGED

Wave 74 audited every consumer and rebuilt them bottom-up as a single slot-configured OITable family, rather than adding one more variant.

WHAT IT MOVED

37 → 1

table consumers now sit on one canonical organism.

WAVE W74 · DECISIONS LOG

06

WAS BROKEN

Layers leaked. Pages inlined spacing, molecules rebuilt atoms, and nobody agreed who owned the gap between two components.

WHAT I CHANGED

A six-layer hierarchy where skipping a layer fails the build, an eight-file promotion contract, and one spacing law: atoms own their insides, never their outsides, the brick wrapper owns the gaps.

WHAT IT MOVED

Spacing became arithmetic instead of judgement: 24 at every canvas edge, 12 between rows, by construction rather than by review.

RULES 82–84

07

WAS BROKEN

Replacing the surface of a live enterprise product usually means one cliff-edge release that users meet on a Monday morning.

WHAT I CHANGED

No big bang. Each change shipped as a numbered wave, scoped, preservation-gated, verified, logged, wrapping Ant Design in OI-prefixed components token by token.

WHAT IT MOVED

100+

waves from W11 to W110 and counting, and no release that users had to relearn the product for.

WAVE NUMBERS FROM THE LOG

09

OUTCOME

Nobody waits for a

designer any more

WHAT THE SYSTEM IS HELD TO, AND WHERE IT STANDS

IT WORKED IF	RESULT	STATUS
No component touches a raw value	Counted in the source: 0 of the 1,304 colour declarations in the system's component stylesheets is a hex, rgb, hsl or named colour	MET
Dark mode costs no component code	Shipped as a token swap; no component code was touched	MET
A sentence becomes a prototype without waiting for a designer	About 15 seconds from plain English to a working Storybook prototype in the production components	MET
A decision outlives the person who made it	Numbered rules, retired ones kept, and an append-only decisions log read by people and agents	MET

3,396
CSS custom properties
COUNTED FROM SOURCE · SEE THE COLOPHON

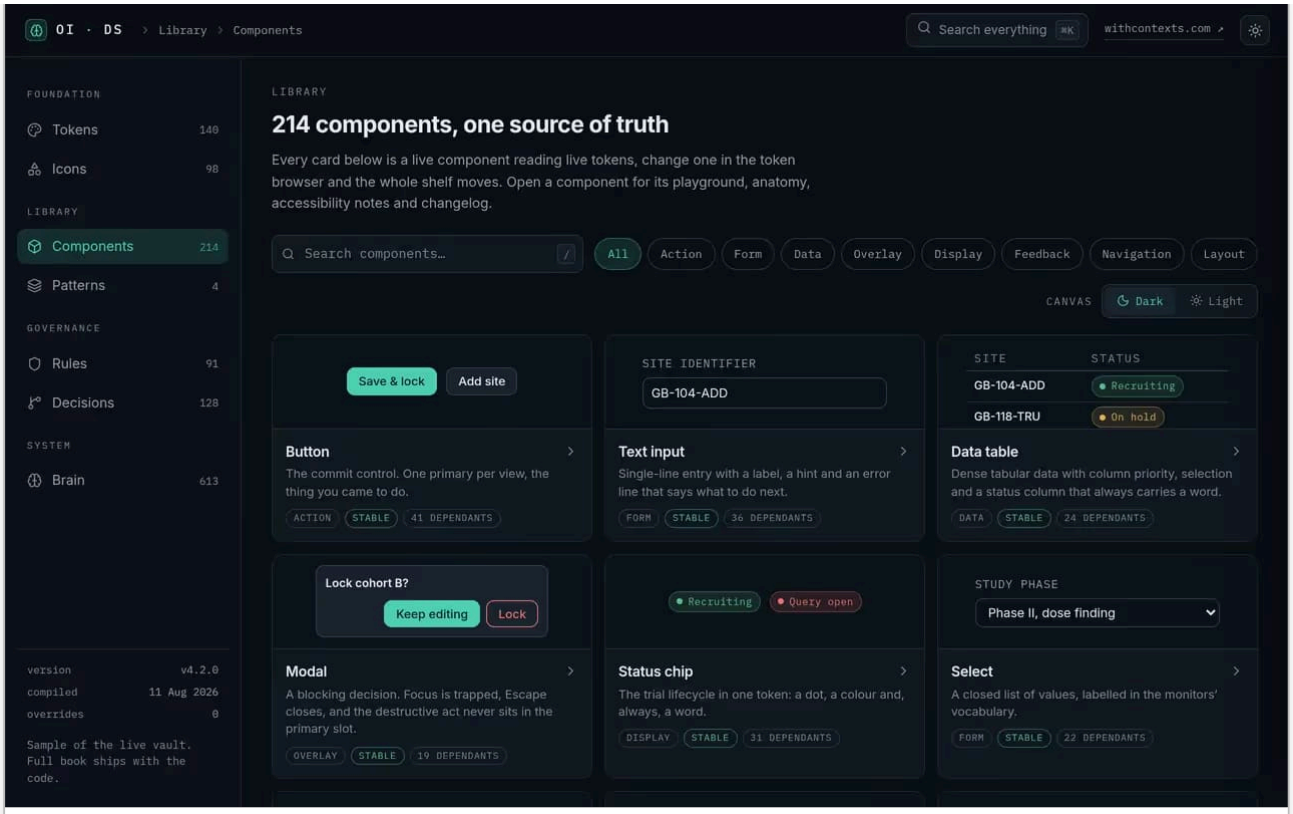
91
numbered rules
TWO RETIRED · NUMBERS KEPT

378 / 6,683
brain documents / cross-links
FROM THE VAULT MANIFEST

~15 s
prototype iteration loop
PLAIN ENGLISH · STORYBOOK OUT

Every figure above is one I can count from the system itself and show you the command for.

The measure of the system is who uses it. Nobody waits for a designer to try an idea any more, a PM with a sentence of plain English gets a working prototype in the product's own components, and the design conversation moved from pixels to rules: when we disagree now, we cite a number, look up why it exists, and either follow it or retire it in the log.



Each component with its status, its version and what depends on it, as the library lists them today.

What I would do differently

01 **Win the engineers before the library.** The system was not the hard part; getting engineers to want it was. I would put the rule book and the memory in front of them before the first component, not after.

02 **Read the architecture before writing a rule.** I once mandated a gutter across the system, hardened it into a rule and a guard, and reverted all three hours later: bricks carry their own padding because they are the frame. The decision log keeps the entry so nobody repeats it.

03 **Compare the options before building one.** The prototyping route that was killed was already under construction when five options were finally compared in a written memo. The memo should have come first.

SEE ALSO

002 · Astrato

Specs shipped as designed: the handoff a system makes cheap, across two pods and eighteen engineers.

003 · Safe Space

Rules decided by refusal: a privacy model settled the way this system's rules are, case by case.

SOURCE NOTES

Every figure on this page is counted from the design-system source, which lives in the same repository as this site: 3,396 distinct --oi-* custom properties and 4,255 declarations of them, 231 components across the three live layers, 23 frozen legacy and 14 templates. The counting method, the command that reproduces each number and the date the source was last touched are published in full on the colophon.

Those numbers used to read 216, 220 and 3,491, depending on which part of this site you were standing in. They were each true of a different count taken at a different time, and nothing was watching, so they drifted apart in public. That is fixed in the way a design system fixes things: one canonical source, and a test that fails the build when a page disagrees with it.

The token count then read 3,444 until v7.38. It had a command, but the command counted the token docs and story files as well as the stylesheets, and nothing ever ran it again. Now a suite runs every command before each release and fails if one prints a different number.

Brain figures (378 documents, 6,683 cross-links, 17 sections) computed from the vault manifest.

Wave and rule numbers (W11–W110+, 91 numbers, 89 live rules, two retired) from the append-only decisions log.

The artifacts are served in full on this domain, the story site, the deck and the class, embedded above.

← PREVIOUS

JBS · the spread, and what it is made of

NEXT →

Astrato · three years inside a data platform

Shad Abbaslou
Product Design Lead · London
withcontexts.com
Last updated 11 Nov 2026

LinkedIn ↗ Read the CV