



Bioconductor's web and package infrastructure

Sustainable open research infrastructure: lower cost, less to run, faster to change

Sean Davis

BIOCONDUCTOR TECHNICAL ADVISORY BOARD

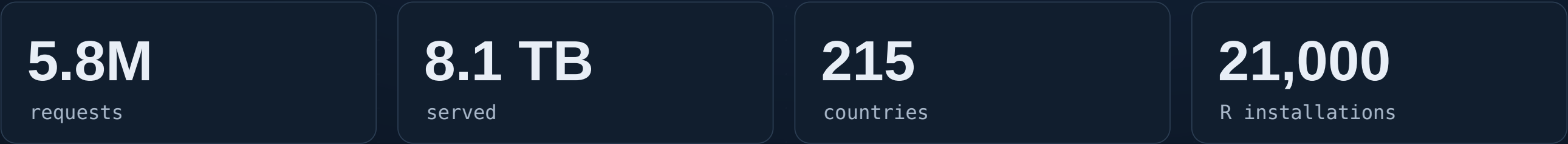
2026-10-01

Where we are, in one slide

- **bioconductor.org has been served by the new stack since 2026-09-28.** It's a Cloudflare Worker in front of R2 object storage. `BiocManager::install()` works unchanged: 8 of 8 acceptance checks pass for release 3.23 and devel 3.24.
- **Why we moved:** crawlers overwhelmed one EBS-backed VM, and nobody could see them.
- **Cost:** about \$5,000 a month of AWS is being retired. The new stack runs at roughly \$240–500 a month (an estimate at list prices).
- **Packages** are still built by BBS and r-universe, and copied hourly from master. The registry built on r-universe is running, but it doesn't serve bioconductor.org yet.
- **Packages and the website are independent:** each publishes on its own path, and site changes go through PRs with previews and one-line rollback.

One day on bioconductor.org

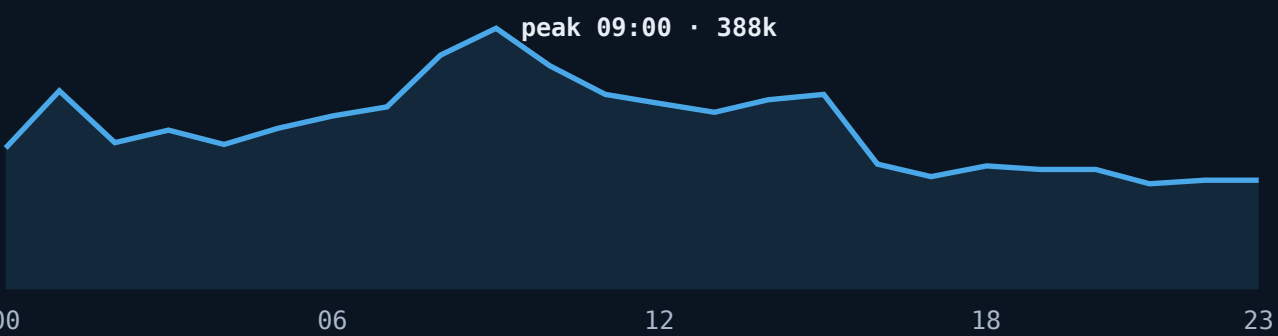
2026-09-30, UTC · FROM THE NEW EDGE'S REQUEST LOGS



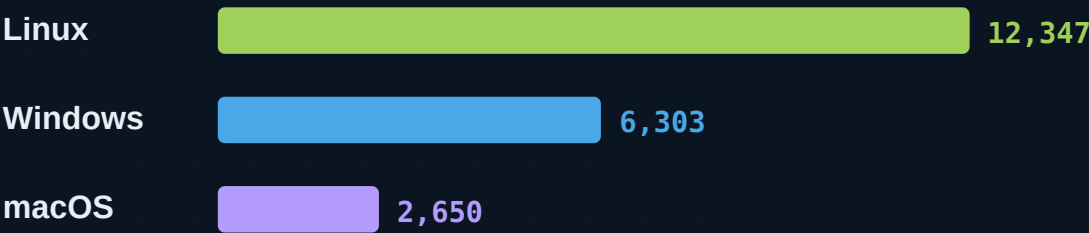
WHO IS ASKING · share of requests



REQUESTS PER HOUR, UTC



R INSTALLATIONS BY OS



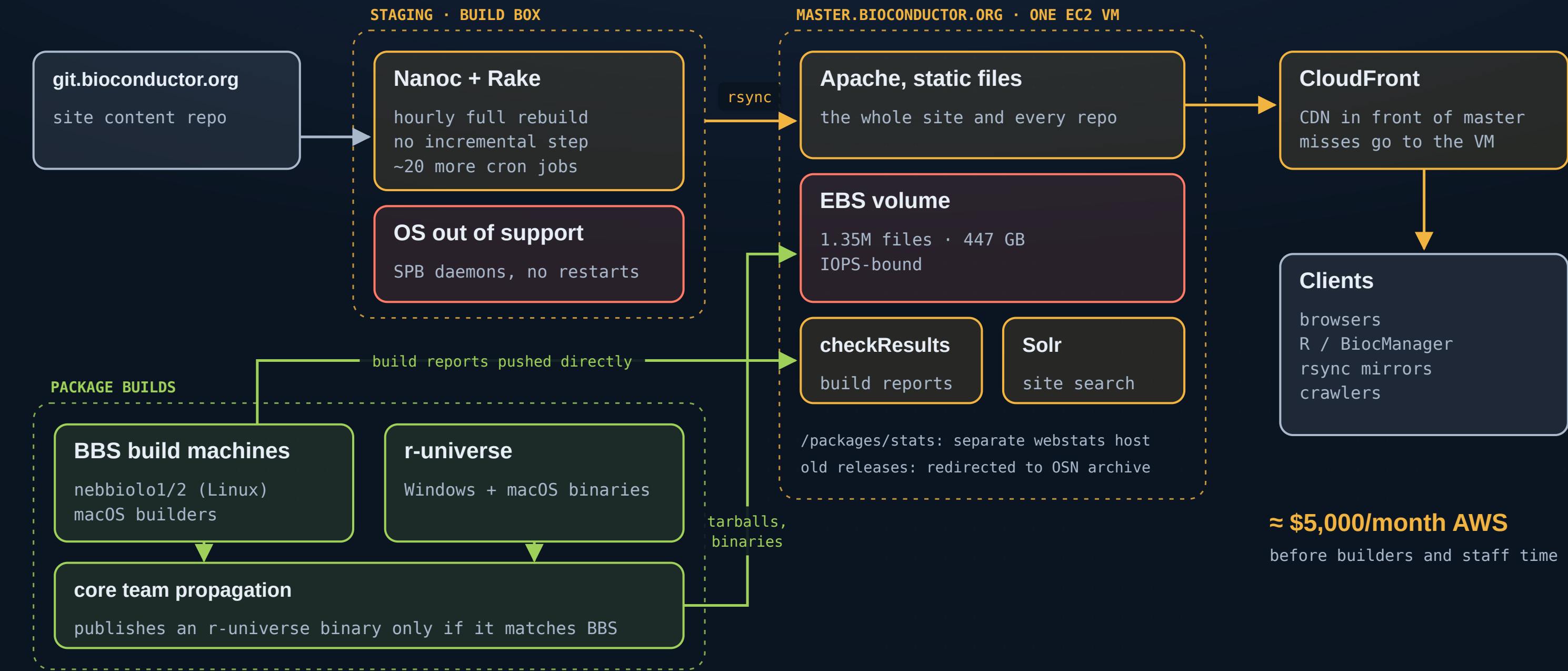
R 4.6 is 56% of them

One machine: 162,163 downloads of Bioconductor 3.19, 40% of R installs that day.

Build reports: 22% of all requests, the site's other front door.

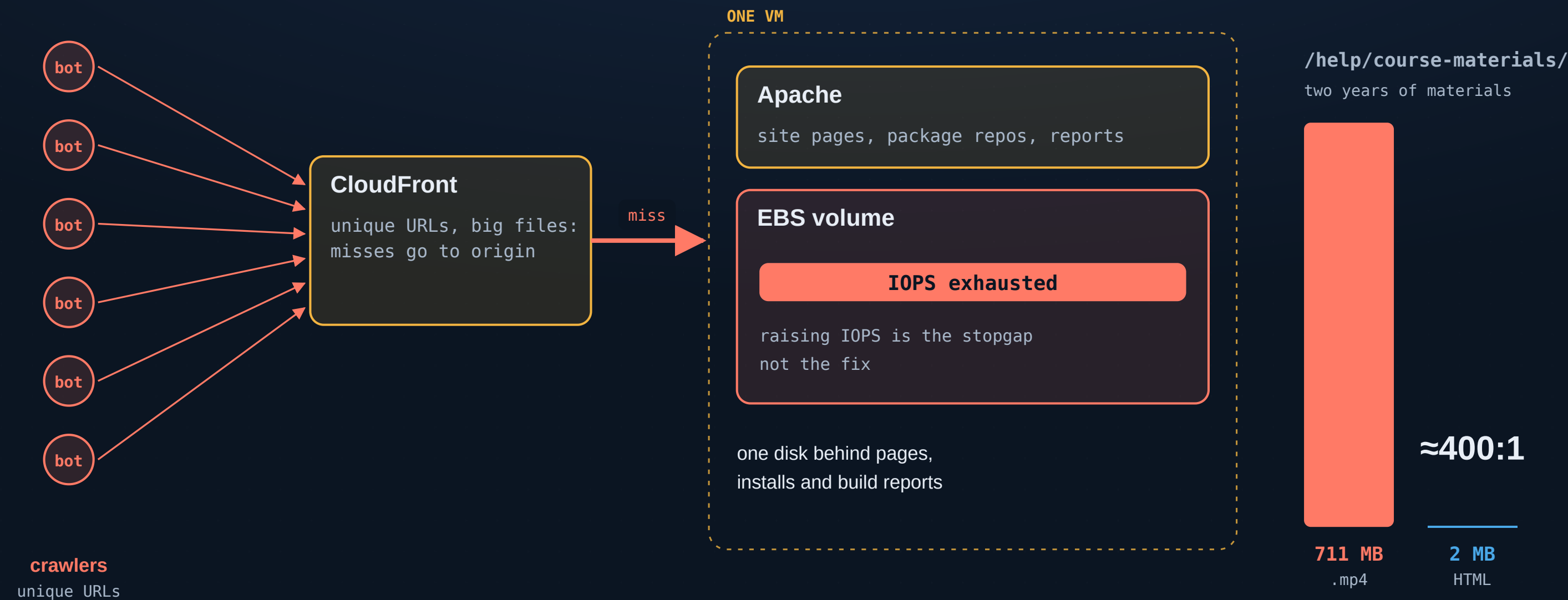
The legacy estate

WHERE WE STARTED



A package repository with a website attached: every `install.packages()`, every mirror, every build report and every page came off this one VM.

The crisis: crawlers against one disk



Why not just fix the VM?

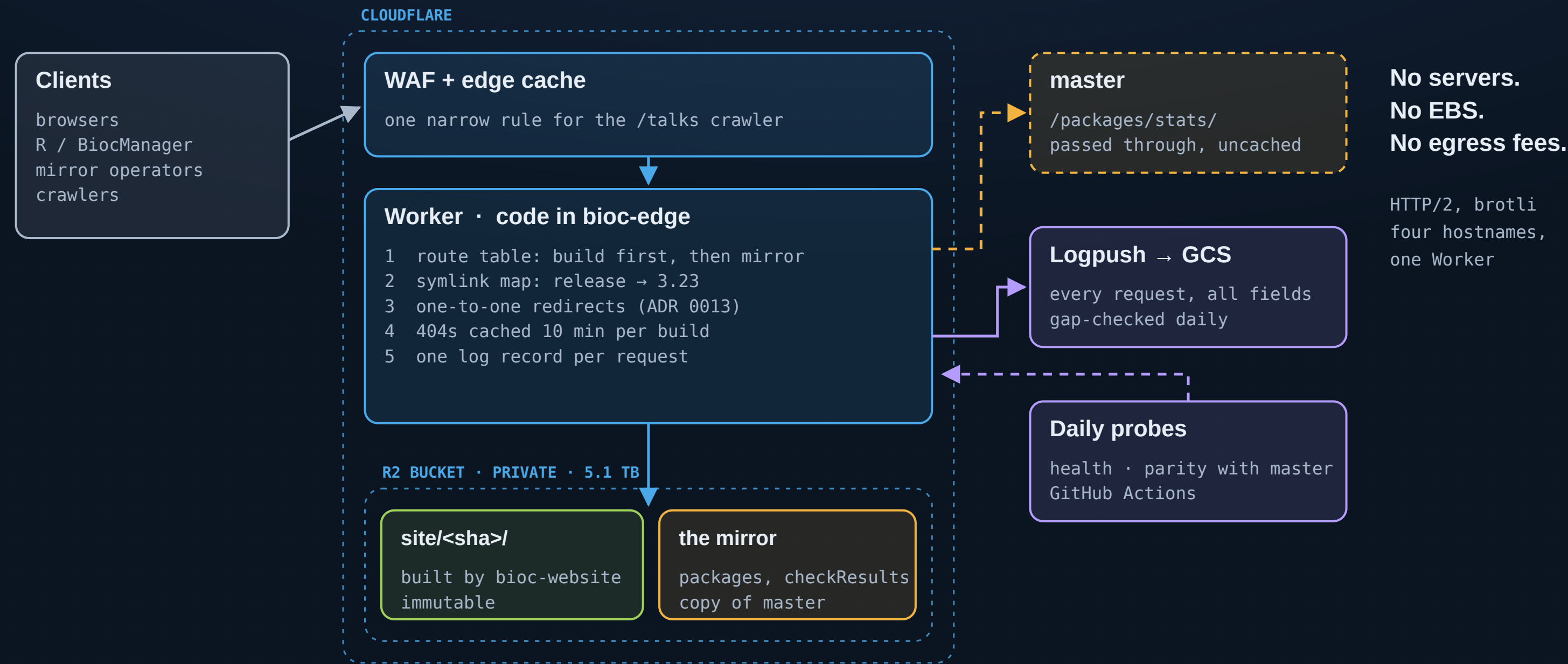
- **One VM, one disk, behind everything.** Pages, installs, mirrors and build reports share one failure domain.
- **Nobody could see the traffic.** Request logs fed statistics jobs; nobody read them request by request. The download statistics have never filtered bots: the code that would has been commented out.
- **The build box is fragile.** Its OS is out of support, it does a full rebuild every hour, and its daemons run with no supervisor.
- **Most of the ~\$5,000 a month is egress,** and sending tarballs to the world is the whole job.
- **Nothing was written down** in a place a newcomer could find it.

i The decision

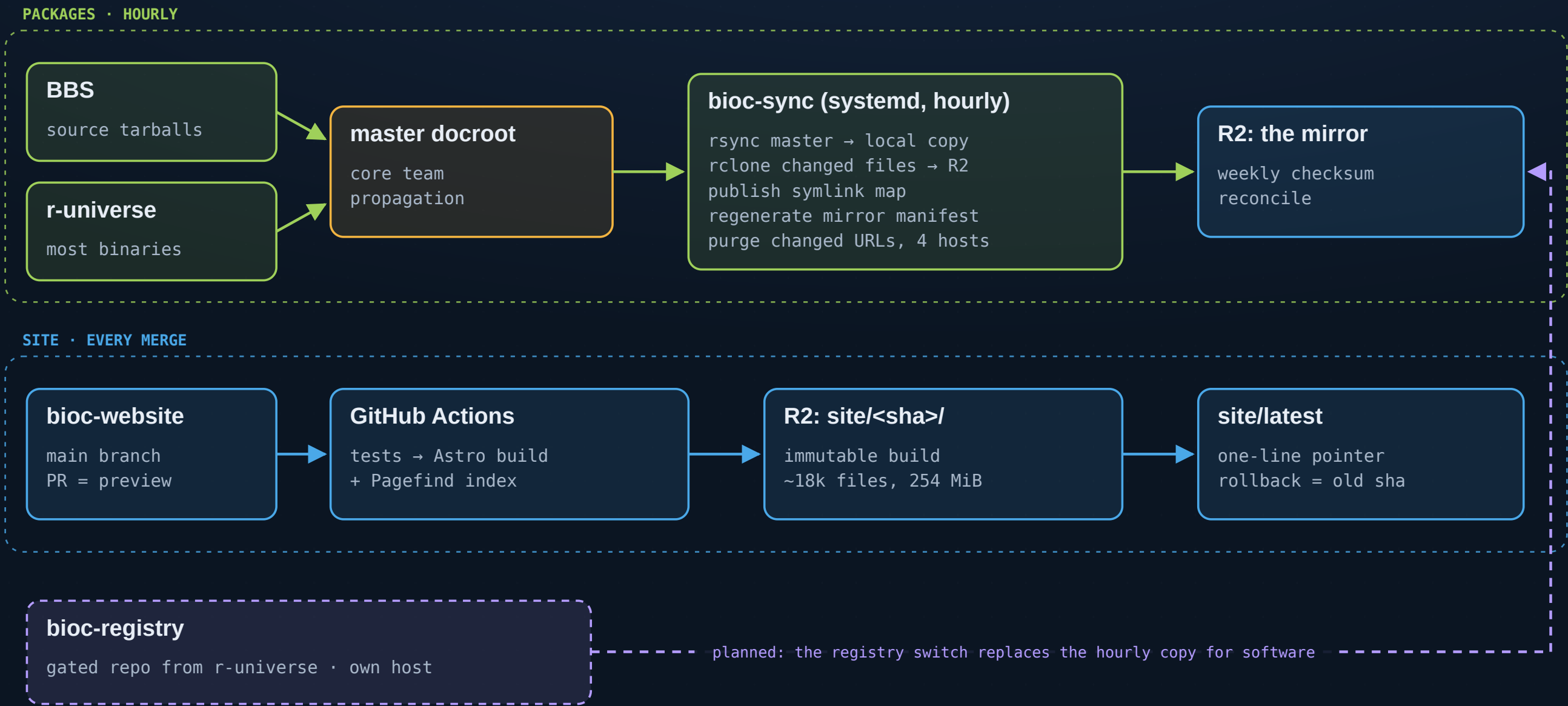
Move serving to object storage plus edge compute, and replace the legacy pieces one at a time behind the live site (a strangler migration), with every choice recorded as an ADR.

Serving: how a request is answered

THE NEW STACK

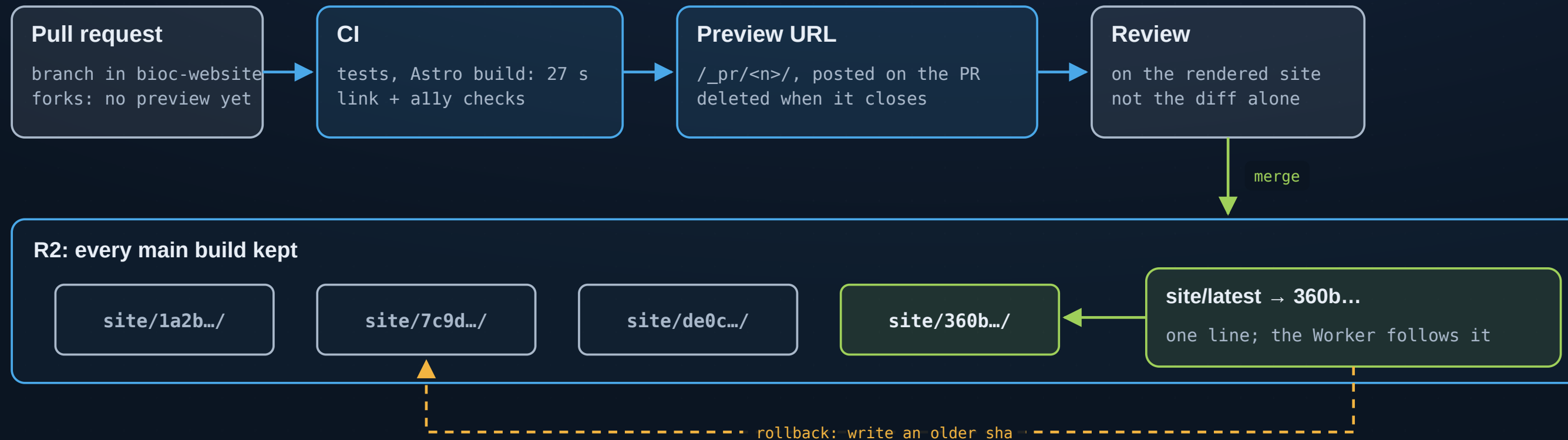


Publishing: how content gets there



Packages and the website are independent: neither waits for the other; they meet only in R2.

Content management: the website as a PR workflow



the response header x-bioc-build names the build that answered

- **Every change is a PR**, with a preview
- **Every build is kept**; rollback is one line
- **Astro build: 27 s**; upload: ~18 min
- **Prose**: still from upstream, by snapshot

Restyling: bioconductor.org/next/

Bioconductor

OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

About

Learn

Packages

Developers

Funding

Donate

Search

Get Started

Home

Bioconductor 3.23

Software

limma

limma

Linear Models for Microarray and Omics Data

Bioconductor version: 3.23 · Package version: 3.68.5

Data analysis, linear models and differential expression for omics data.

Installation

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("limma")
```

Details

Maintainer	Gordon Smyth <smyth@wehi.edu.au>
Author	Gordon Smyth [cre, aut] (ORCID: <https://orcid.org/0000-0001-9221-2892>), Yifang Hu [ctb], Matthew Ritchie [ctb], Jeremy Silver [ctb], James Wettenhall [ctb], Davis McCarthy [ctb], Di Wu [ctb], Wei Shi [ctb], Belinda Phipson [ctb], Aaron Lun [ctb], Natalie Thorne [ctb], Alicia Oshlack [ctb], Carolyn de Graaf [ctb], Yunshun Chen [ctb], Goknur Giner [ctb], Mette Langaas [ctb], Egil Ferkingstad [ctb], Marcus Davy [ctb], Francois Pepin [ctb], Dongseok Choi [ctb], Charity Law [ctb], Mengbo Li [ctb], Lizhong Chen [ctb] (ORCID: <https://orcid.org/0000-0002-8326-6781>)

Bioconductor

NEXT PREVIEW

next

bioc-release

limma

limma

Linear Models for Microarray and Omics Data

v3.68.5

checks: error

propagated 2026-08-10

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("limma")
```

ABOUT

Data analysis, linear models and differential expression for omics data.

DOCUMENTATION

Reference manual

A brief introduction to limma

limma User's Guide

BUILD & CHECK STATUS

bioc-checks ERROR

linux-devel-arm64 NOTE

linux-devel-x86_64 NOTE

linux-release-arm64 NOTE

linux-release-x86_64 NOTE

macos-oldrel-arm64 NOTE

macos-oldrel-x86_64 OK

macos-release-arm64 NOTE

macos-release-x86_64 OK

source OK

wasm-release OK

windows-devel-arm64 NOTE

windows-devel-x86_64 OK

windows-oldrel-x86_64 OK

windows-release-arm64 NOTE

windows-release-x86_64 OK

MAINTAINER

Gordon Smyth <smyth@wehi.edu.au>

AUTHORS

Gordon Smyth [cre, aut] (ORCID: <https://orcid.org/0000-0001-9221-2892>), Yifang Hu [ctb], Matthew Ritchie [ctb], Jeremy Silver [ctb], James Wettenhall [ctb], Davis McCarthy [ctb], Di Wu [ctb], Wei Shi [ctb], Belinda Phipson [ctb], Aaron Lun [ctb], Natalie Thorne [ctb], Alicia Oshlack [ctb], Carolyn de Graaf [ctb], Yunshun Chen [ctb], Goknur Giner [ctb], Mette Langaas [ctb], Egil Ferkingstad [ctb], Marcus Davy [ctb], Francois Pepin [ctb], Dongseok Choi [ctb], Charity Law [ctb], Mengbo Li [ctb], Lizhong Chen [ctb] (ORCID: <https://orcid.org/0000-0002-8326-6781>)

Today: from BBS, via master’s repositories

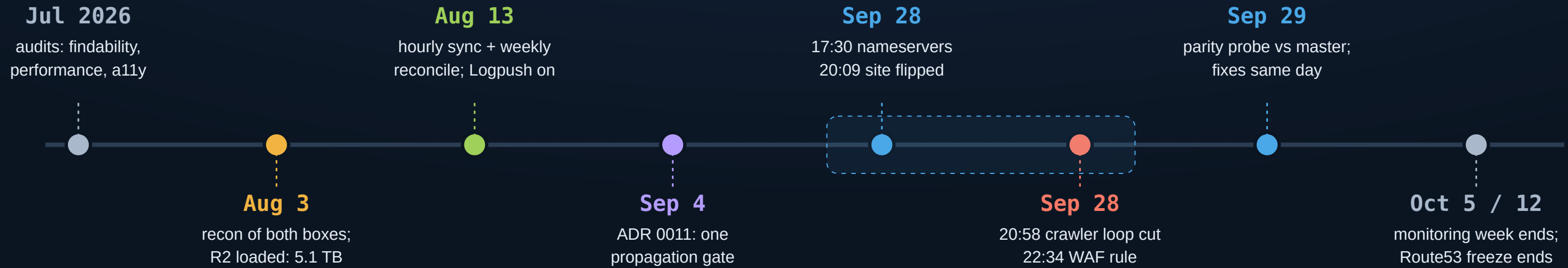
/next/: from the bioc-registry data plane

Same package, **different data path**: `/next/` renders from r-universe builds through the registry’s gate, with no dependency on the legacy AWS servers. New designs can be tried there in public without touching the live pages. [Browse it](#)

Bioconductor infrastructure · Technical Advisory Board

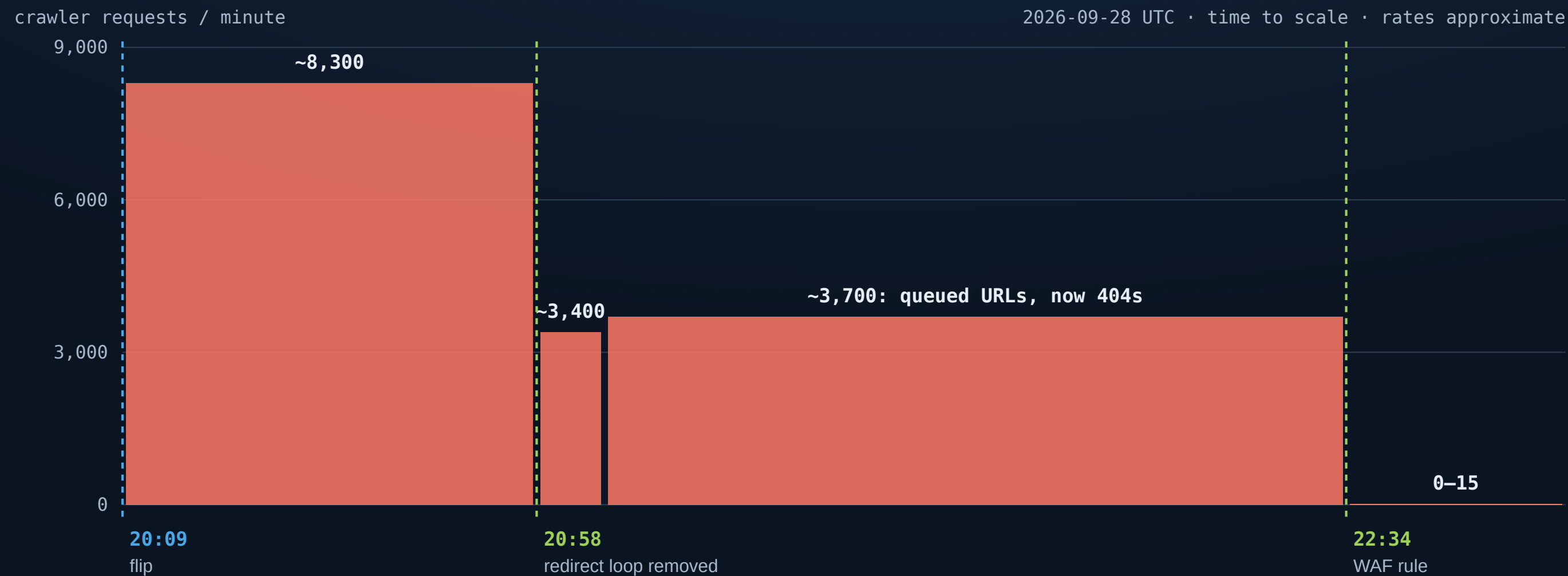
10

The cutover



- `BiocManager::install()` acceptance passed on release and devel **after** the flip
- Rollback stays one step away: Route53 is kept, with records frozen, until 2026-10-12

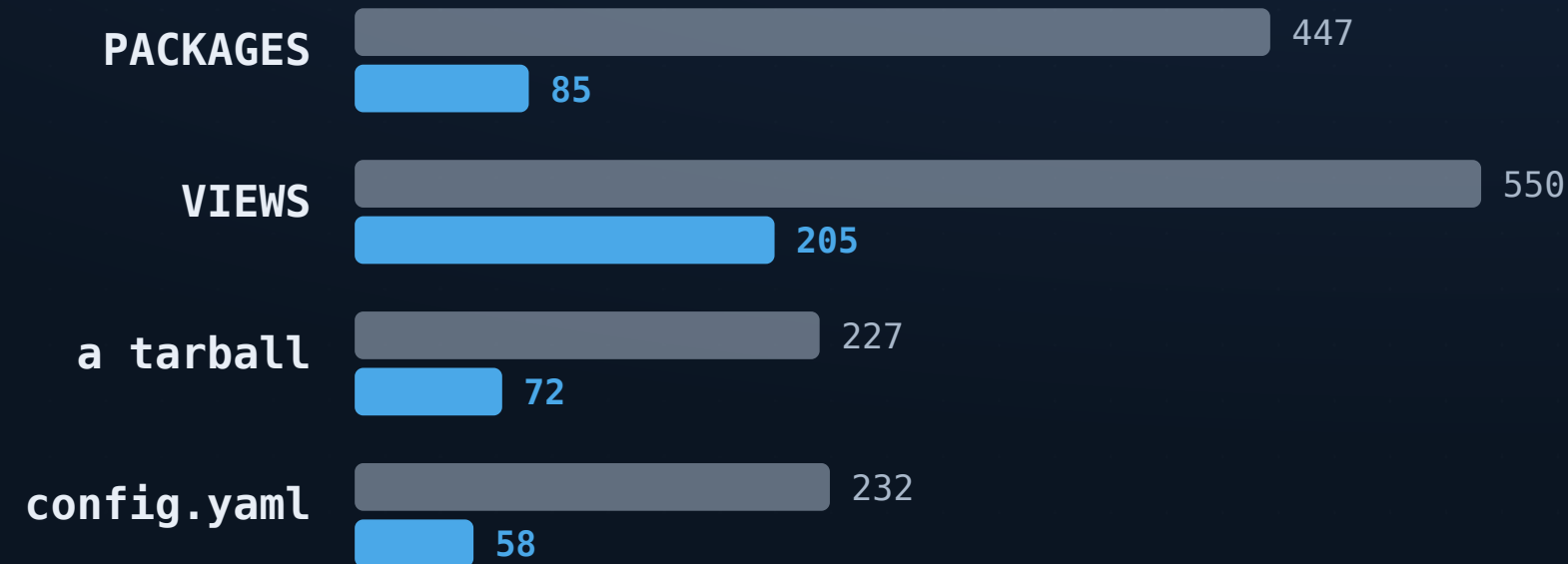
Day one: three quarters of the traffic was one crawler



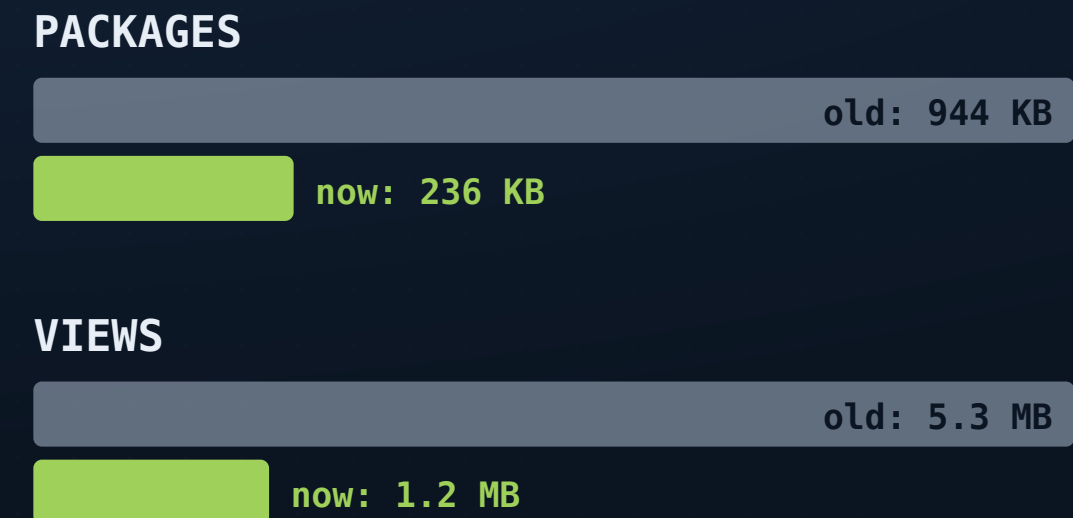
84,414 of the first 112,632 requests came from 1,130 addresses on two cloud networks, disguised as Mac Chrome. They fed a redirect loop: `/talks/*` sent everything to an index page full of relative links.

Faster, smaller downloads

time to first byte from Denver, ms (before → after the cache fix)



bytes on the wire, compressed



- HTTP/2 and brotli everywhere. The old site spoke HTTP/1.1 and sent the indexes uncompressed.
- 63% of 530,455 requests served from the edge cache (two hours on 2026-09-29)
- Still open: CloudFront was 30–50 ms faster from this host; part of that is Worker time

Costs

USD per month



Cloudflare list prices applied to measured traffic, not a bill. AWS figure excludes builder hardware and staff time.

What the parity probe found

We compared 13,271 paths from real traffic against master. Fixed the same day:

- 404s returned 9 bytes of text instead of the site's 404 page
- Package files skipped the edge cache
- Double-slash links, range requests past a file's end
- Landing pages had lost their Windows and macOS download links

The remaining edge cases (bioc-edge #43–#47) are small, tested code changes, because the web server is now code: routing rules, redirects and directory listings live in one program, with tests, not in server config.

Where packages come from today

PACKAGES AND R-UNIVERSE

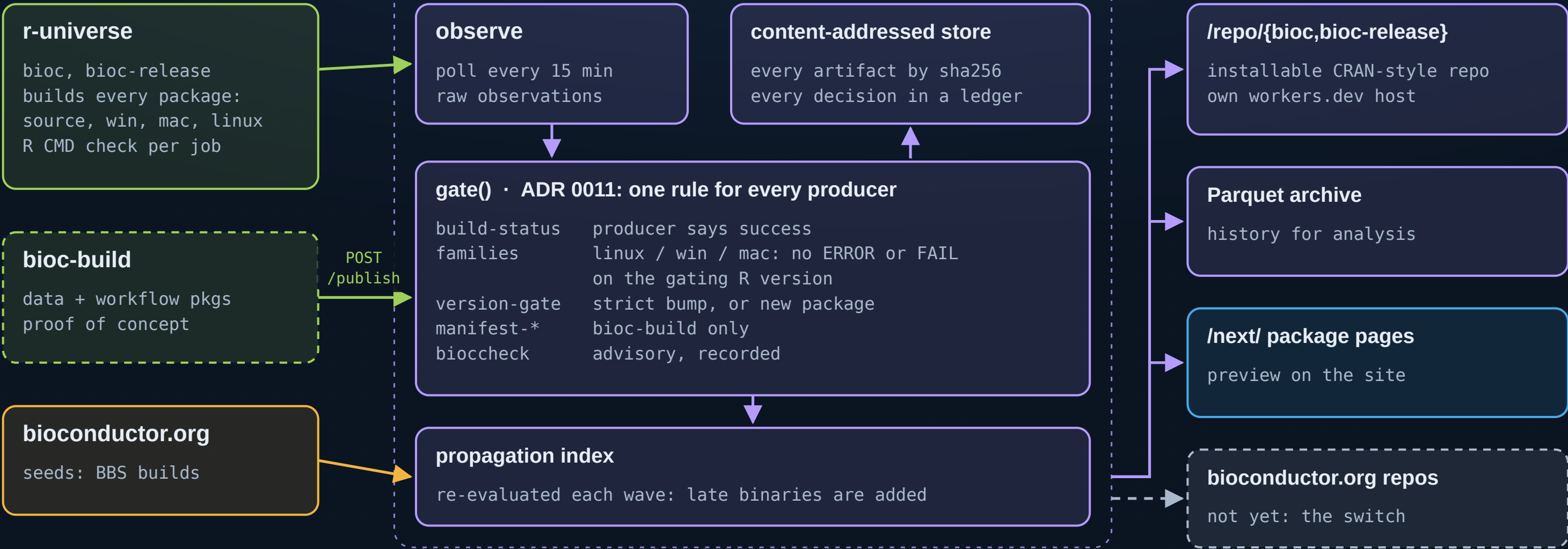
PACKAGE TYPE	BUILT BY	REACHES BIOCONDUCTOR.ORG VIA
Software, source tarballs	BBS	master, then the hourly sync
Software, Windows and macOS binaries	r-universe, published by core team propagation	master, then the hourly sync
Annotation, experiment data, workflows	legacy builders; no replacement running yet	master, then the hourly sync

r-universe already ships most binaries

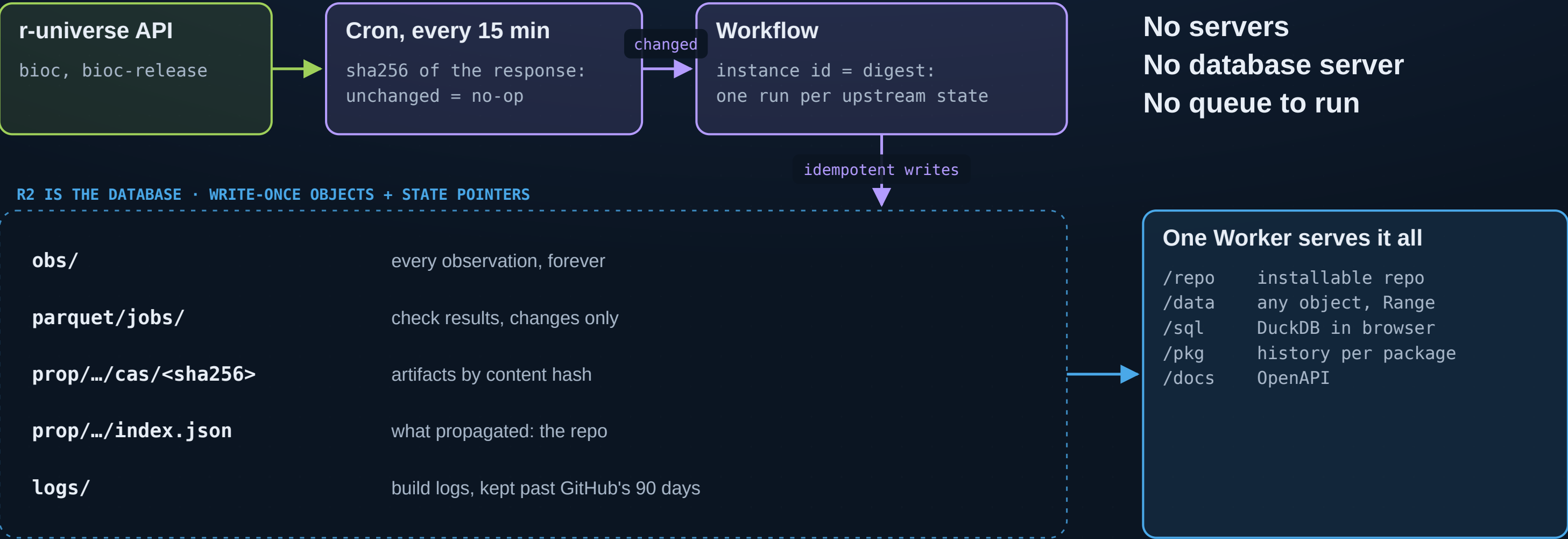
The core team publishes an r-universe binary only when its version and commit match what BBS built. Across 2,326 packages, r-universe's R CMD check agrees with BBS's nebbio1 96.7% of the time (counting WARNINGS as clean).

bioc-registry: one gate for every producer

BIOC-REGISTRY · CLOUDFLARE WORKER



bioc-registry: serverless, with storage as the database



Using the registry today

```
1 # A preview, not for production
2 reg <- "https://bioc-registry.seandavi.workers.dev/repo/bioc-release"
3 # devel: .../repo/bioc
4
5 install.packages("edgeR",
6   repos = c(BioCsoft = reg,
7             CRAN = "https://cloud.r-project.org"))
```

⚠ Not through BiocManager

`BiocManager::install(site_repository = reg)` adds the registry *next to* bioconductor.org, and R takes the highest version. BiocManager has no setting that points it at the registry instead.

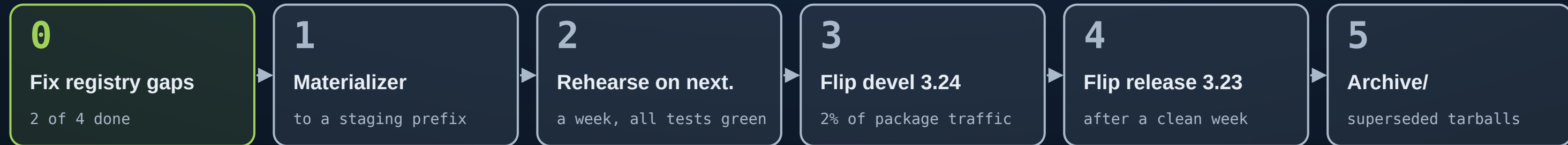
Works (checked 2026-09-30)

- 2,390 release software packages
- Source, Windows and macOS indexes
- A clean install, at the registry's version

Not yet

- `PACKAGES.rds` (R falls back to `.gz`, so installs still work)
- `Archive/` for older versions
- Annotation and experiment-data packages: add those repositories yourself

The registry switch: a plan with tests



each step names its rollback · nothing advances until its tests pass

Compared file by file: 2,374 of 2,384 release source packages are at the same version.

Blocking: no `PACKAGES.rds`, the index R asks for first. Some packages have hard dependencies that can't be installed.

A decision, not a fix: 3 release and 8 devel packages lag. Their newer versions fail in r-universe, so the gate is stricter than BBS here.

Ten regression tests (T1–T10) run on every materialize and block a broken tree.

Data, experiment and workflow packages

434

experiment-data packages, 28 GiB in total

67

of them hold three quarters of the bytes

425 / 433

changed size by under 1% across the last release bump

r-universe builds software packages but has a **100 MiB limit**, and many data packages are far bigger. Until something else builds them, the legacy builders can't be turned off.

bioc-build is a proof of concept:

- a pinned copy of r-universe's own workflow, on GitHub Actions
- a configurable size limit
- packages authorized by a reviewed manifest (bioc-manifest)
- dependencies resolved from the registry, published through the same gate

The 25 workflow packages fit in 319 MiB together.

Roadmap

WHAT'S NEXT

M0

Operational readiness

public repos, CI, badges
daily health probes

mostly done

→ deploys still from a laptop

M1

Production cutover

flipped 2026-09-28
BiocManager 8/8 on 3.23, 3.24

flipped; closing out

→ parity fixes bioc-edge#43–48

M2

Full disconnect

retire master, staging, BBS
legacy stats, ~\$5k/mo AWS

next

→ registry switch; data-pkg builds

M3

Search + redesigns

Pagefind search ADR
redesigns through PRs

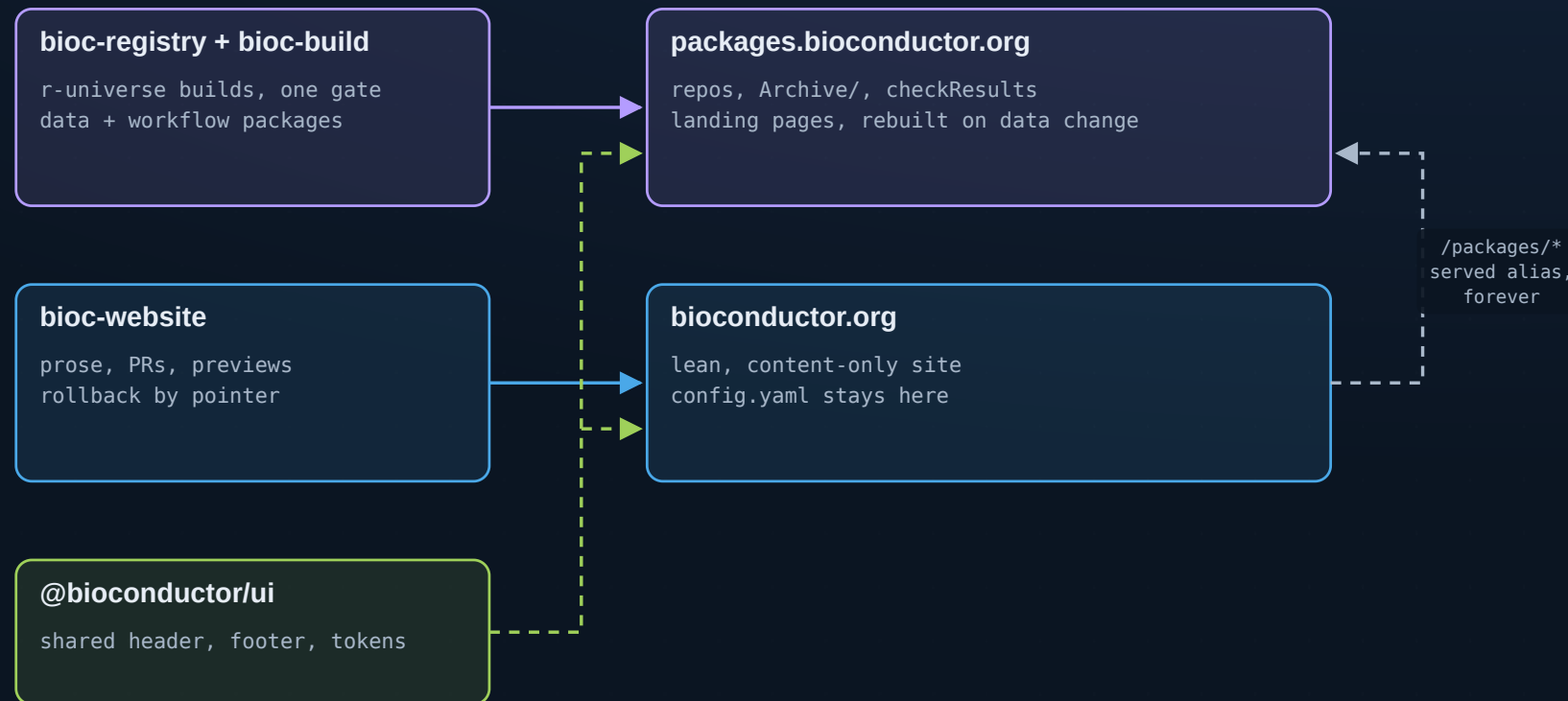
can start now

→ prove the contribution loop

M2 needs: parity with master clean (bioc-edge #43–48) · the registry switch · a build home for data and workflow packages · download statistics from the new logs

Where this could go: two sites

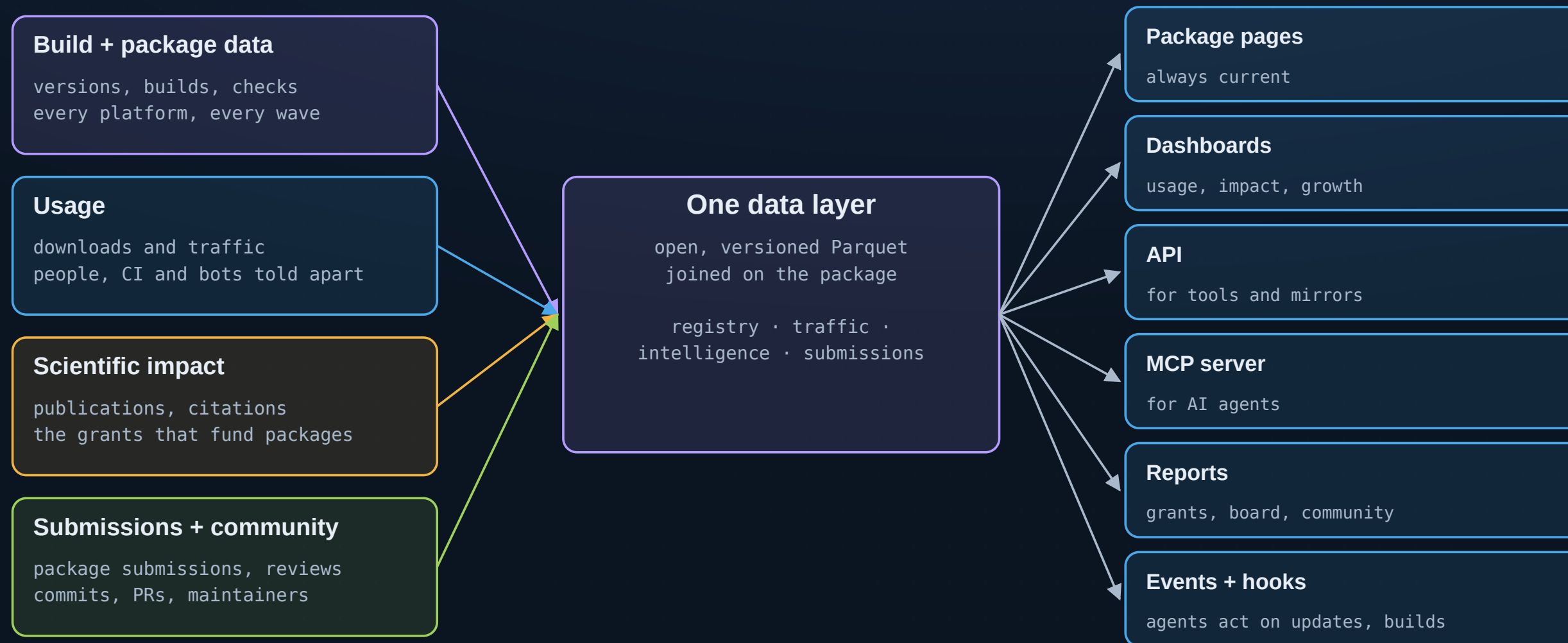
UNDER DISCUSSION, NOT DECIDED



- **packages.bioconductor.org** for everything package-shaped; **bioconductor.org** becomes a lean content site
- **bioconductor.org/packages never breaks:** BiocManager hard-codes it, so it stays as an alias
- **One look across both** from a shared Astro package ([discussion: #71](#))

One data layer for Bioconductor

WHERE THIS COULD GO



Status, usage, impact and community, joined on the package, open to people **and** agents. Explore today:
[bioc-intelligence dashboard](#) · [bioc-registry API docs](#)

Acknowledgements

- **Andres Wokaty** and **Lori Kern**: major help with the website and the r-universe builds
 - **Hervé Pagès**: explaining the build system and the propagation logic
 - **Jeroen Ooms** and **rOpenSci**: r-universe, which most of this stands on
- And the **many people who built and ran Bioconductor's infrastructure over the last 25 years**. Everything here is a migration of their work, and it only worked because that work was there to learn from.

Materials

Docs site: seandavi.github.io/bioc-infrastructure

- [How bioconductor.org works](#)
- [Operations: runbooks, decisions, risks](#)
- [What the new site has changed](#)
- [Roadmap](#)
- [Registry switch plan](#)
- [The propagation gate](#)
- [Architecture diagram](#)
- [Reporting a problem](#)

Explore

- [bioc-registry API docs](#)
- [bioc-intelligence dashboard](#)

Repositories

- [bioc-edge](#): the Worker and the sync
- [bioc-website](#): pipeline and Astro build
- [bioc-registry](#): r-universe observer and gate
- [bioc-build](#) and [bioc-manifest](#): data and workflow packages
- [bioc-infrastructure](#): docs, ADRs, roadmap issues
- [bioc-intelligence](#): impact data and [dashboard](#)

Decisions: the ADRs