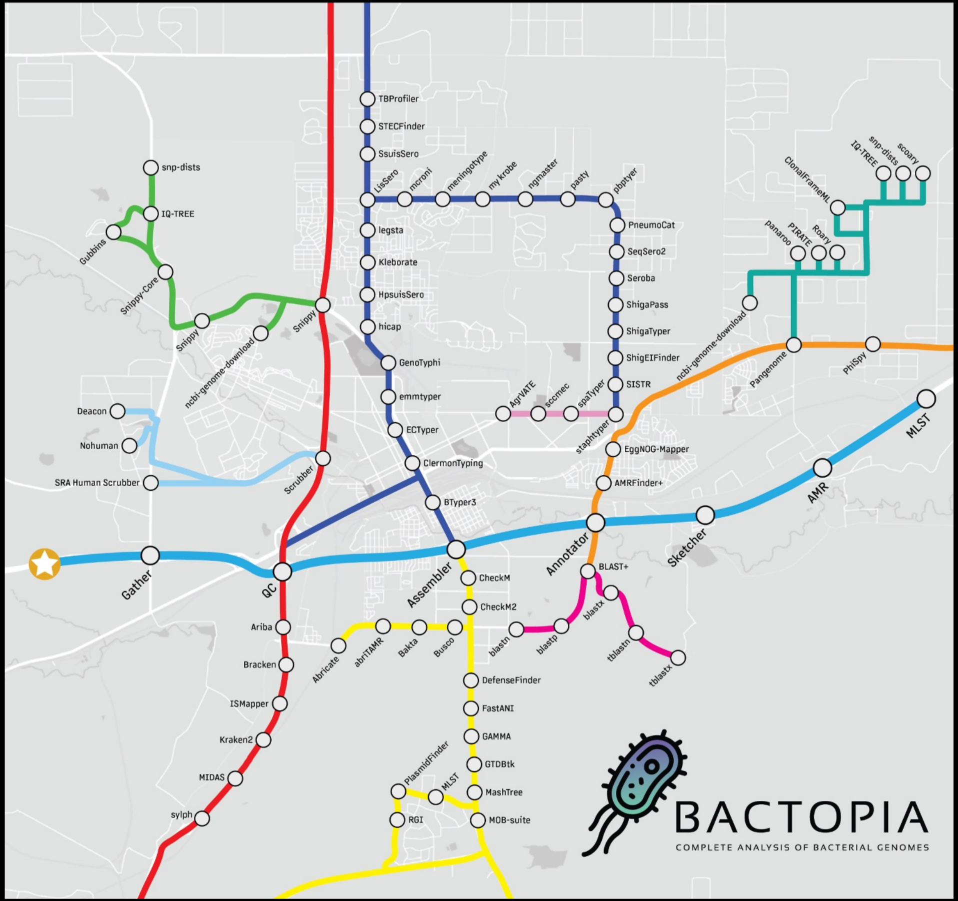
nextflow
SUMMIT
by seqera

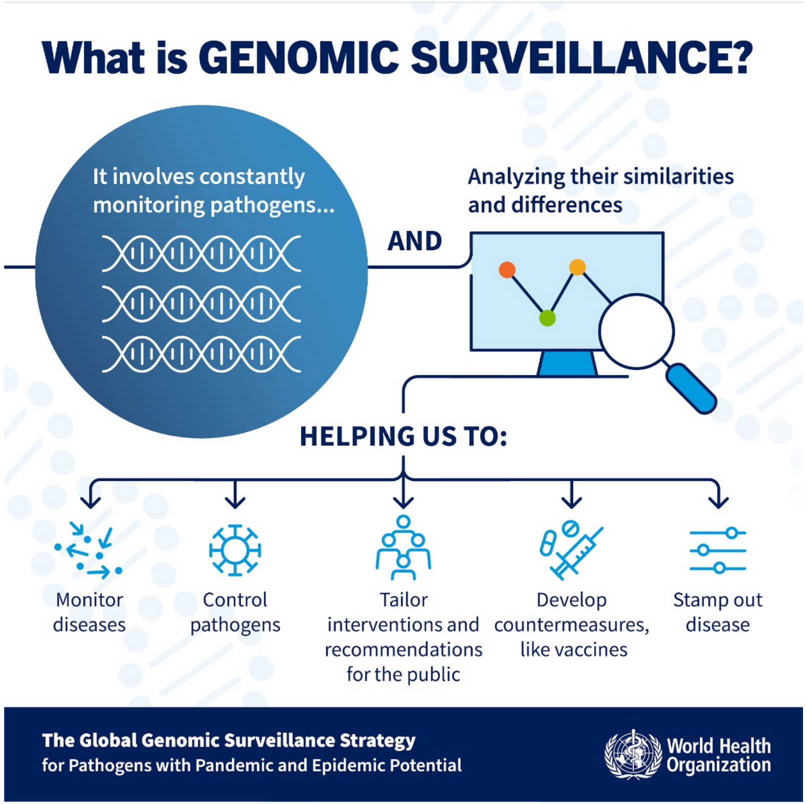
How We Avoided AI Slop in Bactopia

Robert A. Petit III, PhD
Senior Bioinformatics Scientist
Wyoming Public Health Laboratory



Disclaimer, this is a reflection of my experience (and the mistakes I made along the way!) using AI to develop Bactopia v4

Genomic surveillance at Public Health Laboratories (PHL)



What is GENOMIC SURVEILLANCE?

It involves constantly monitoring pathogens...

AND

Analyzing their similarities and differences

HELPING US TO:

- Monitor diseases
- Control pathogens
- Tailor interventions and recommendations for the public
- Develop countermeasures, like vaccines
- Stamp out disease

The Global Genomic Surveillance Strategy
for Pathogens with Pandemic and Epidemic Potential

Source: https://www.who.int/initiatives/genomic-surveillance-strategy

- Many PHLs are now equipped with:
  - Sequencing instruments
  - In-house experts to sequence
- Many PHLs are ***not*** equipped with:
  - In-house bioinformaticians
  - Technical expertise to run pipelines
  - Necessary compute resources

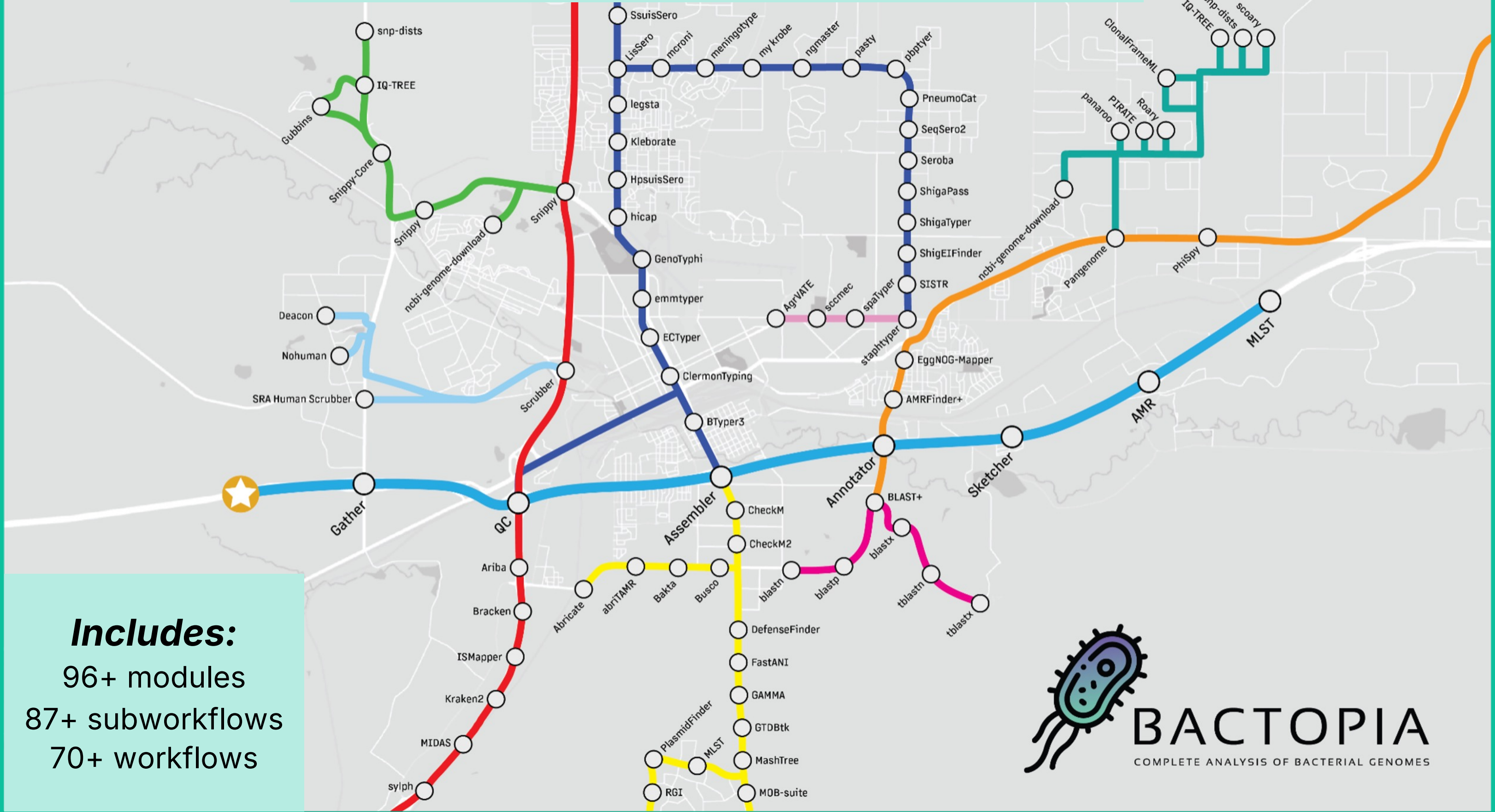
Bactopia provides a low barrier of entry

- End-to-end Nextflow pipeline for bacterial genome analysis
- Easy installation via Conda
- Minimal compute requirements
- Consistently developed and maintained for 7+ years

BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

Bactopia's *“end-to-end”* pipeline



Includes:

96+ modules

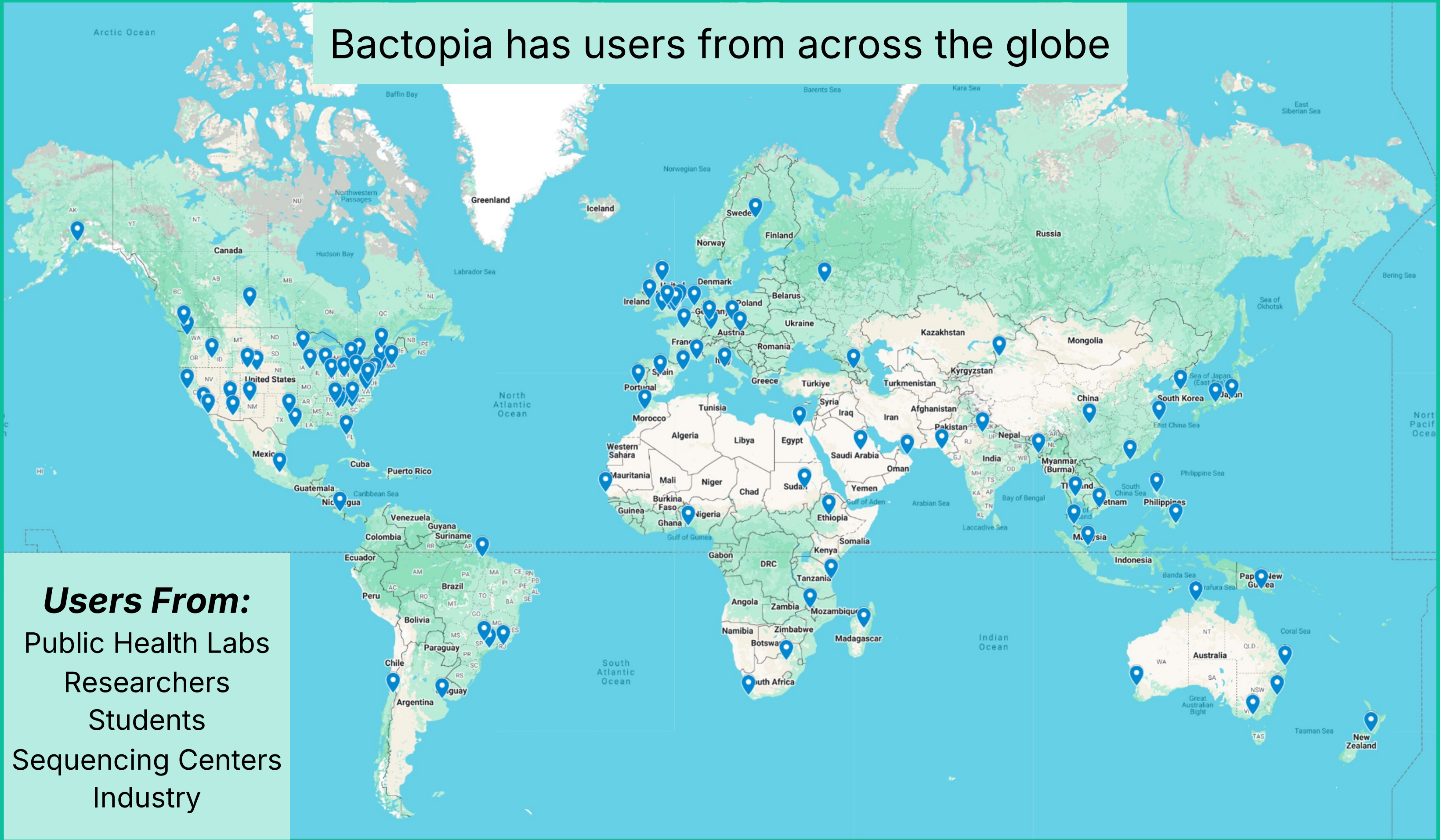
87+ subworkflows

70+ workflows

BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

Bactopia has users from across the globe



Users From:

Public Health Labs
Researchers
Students
Sequencing Centers
Industry

“How many people are on the Bactopia Team?”

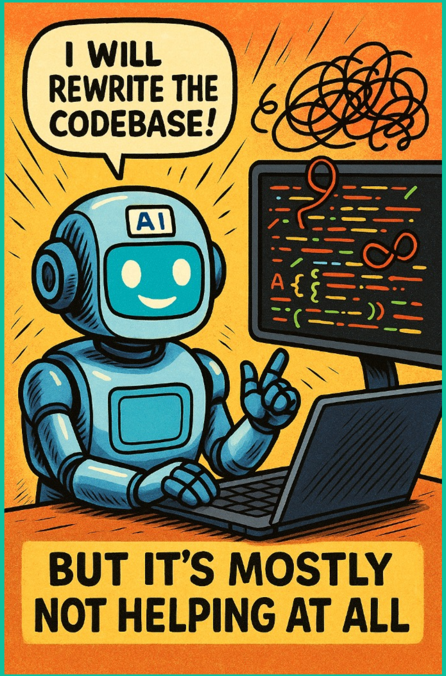
...Just me...

“Oh, so you have a ‘hit by bus’ problem?”

- Random Conversation -

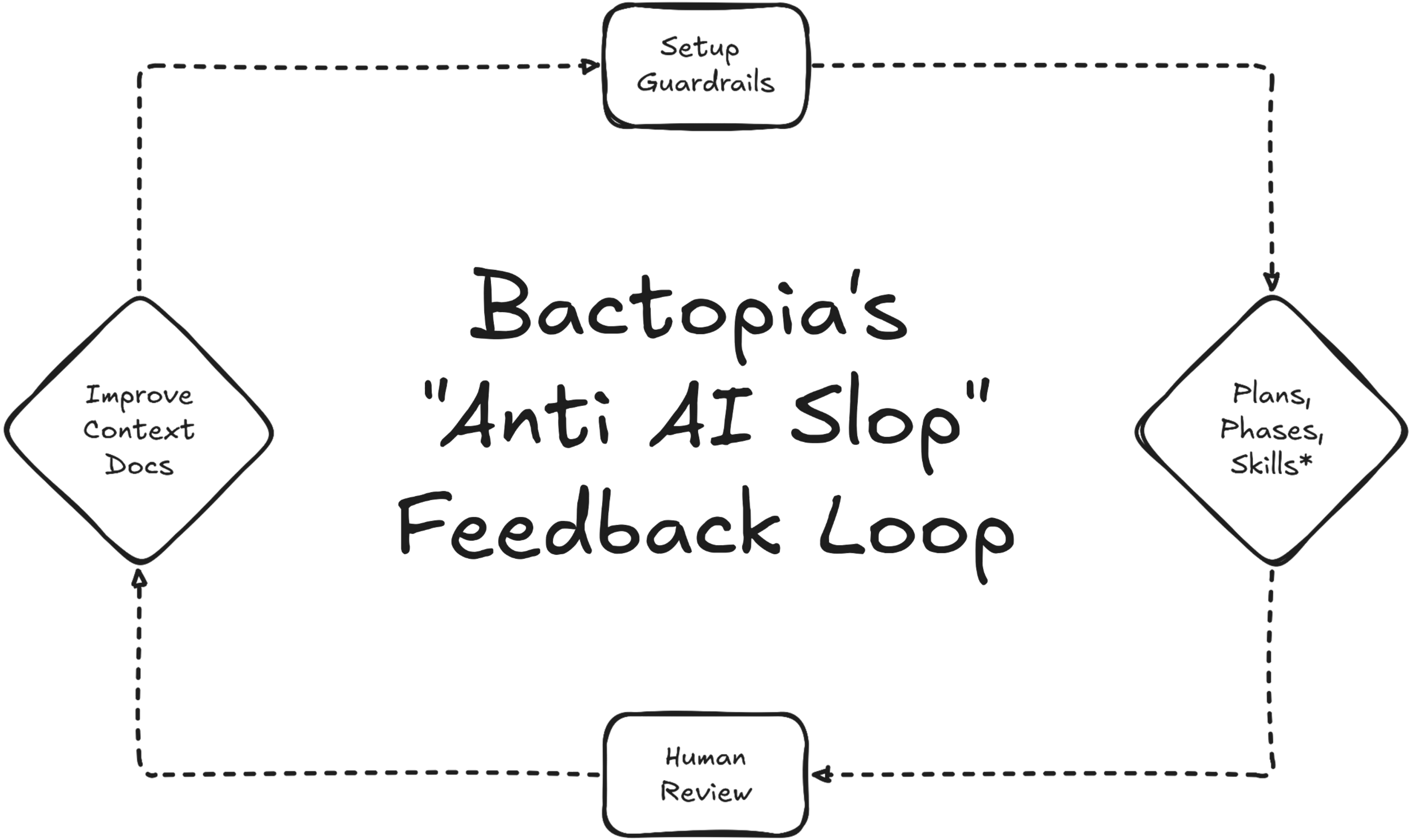
AI should be able to help...

In 2025, AI failed to make a significant impact in the conversion of Bactopia to Nextflow's strict syntax



Back to the drawing board...

Bactopia's "Anti AI Slop" Feedback Loop

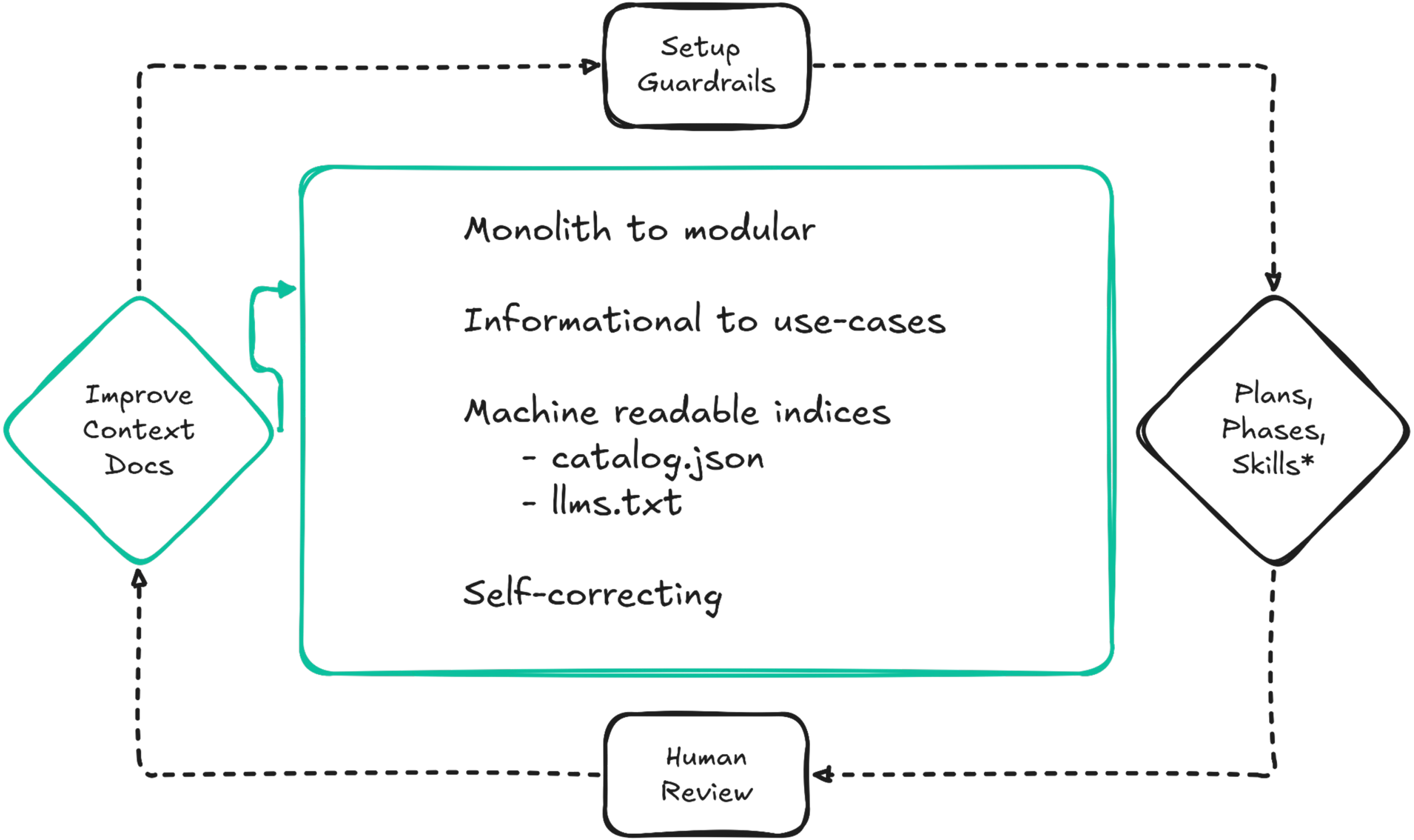


Setup Guardrails

Plans, Phases, Skills*

Human Review

Improve Context Docs



Setup Guardrails

Plans, Phases, Skills*

Human Review

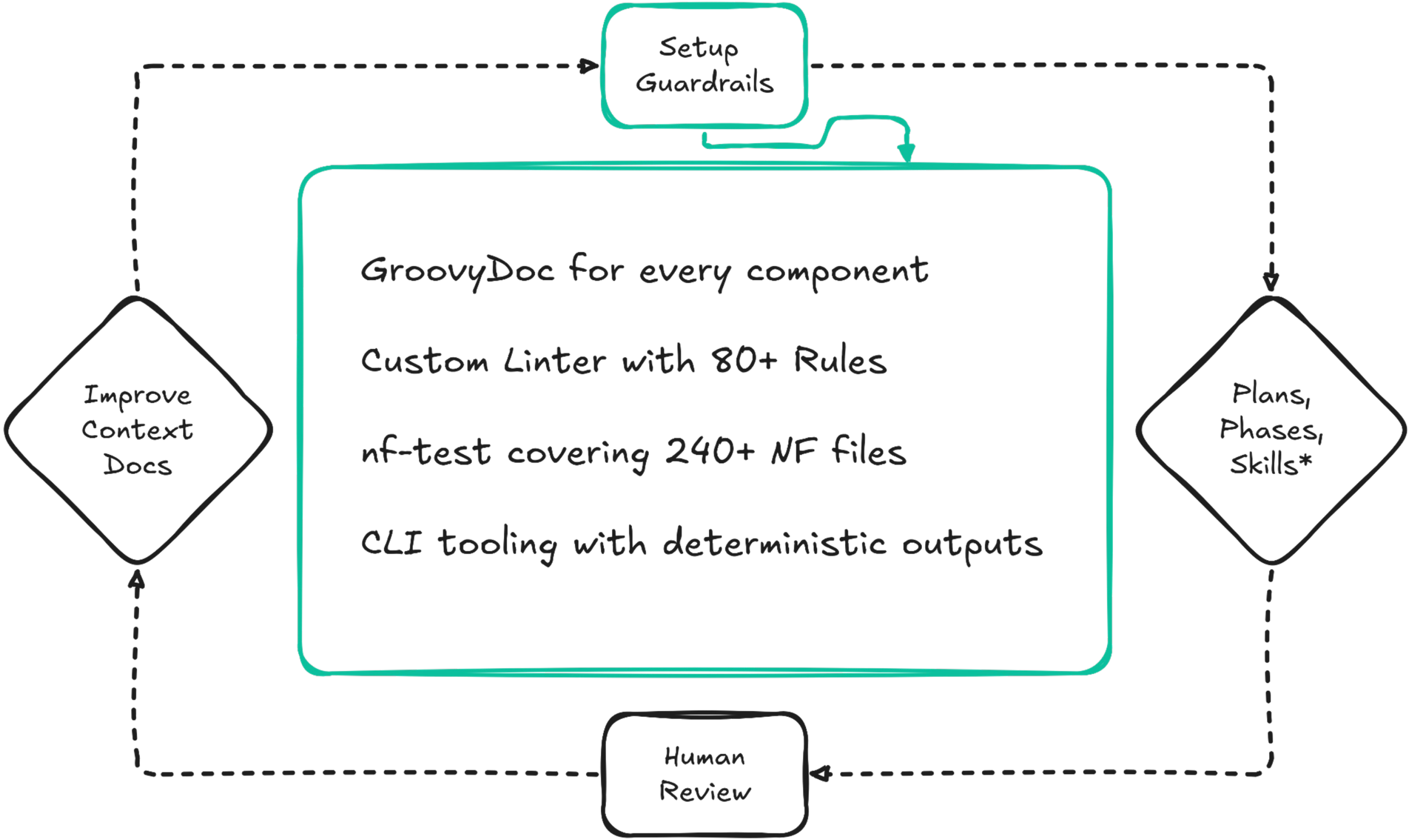
Improve Context Docs

Monolith to modular

Informational to use-cases

Machine readable indices
  - catalog.json
  - llms.txt

Self-correcting

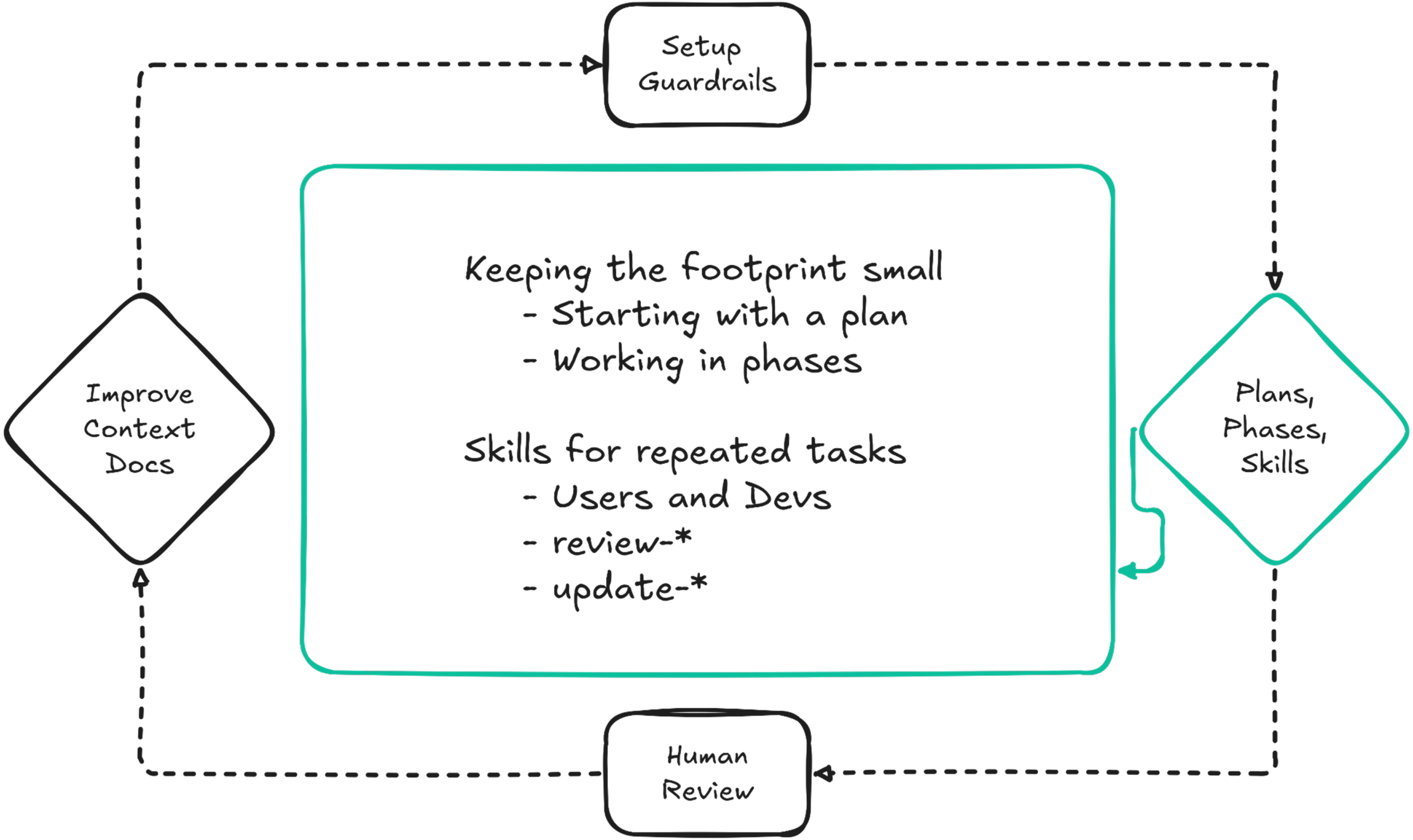


GroovyDoc for every component

Custom Linter with 80+ Rules

nf-test covering 240+ NF files

CLI tooling with deterministic outputs



Setup Guardrails

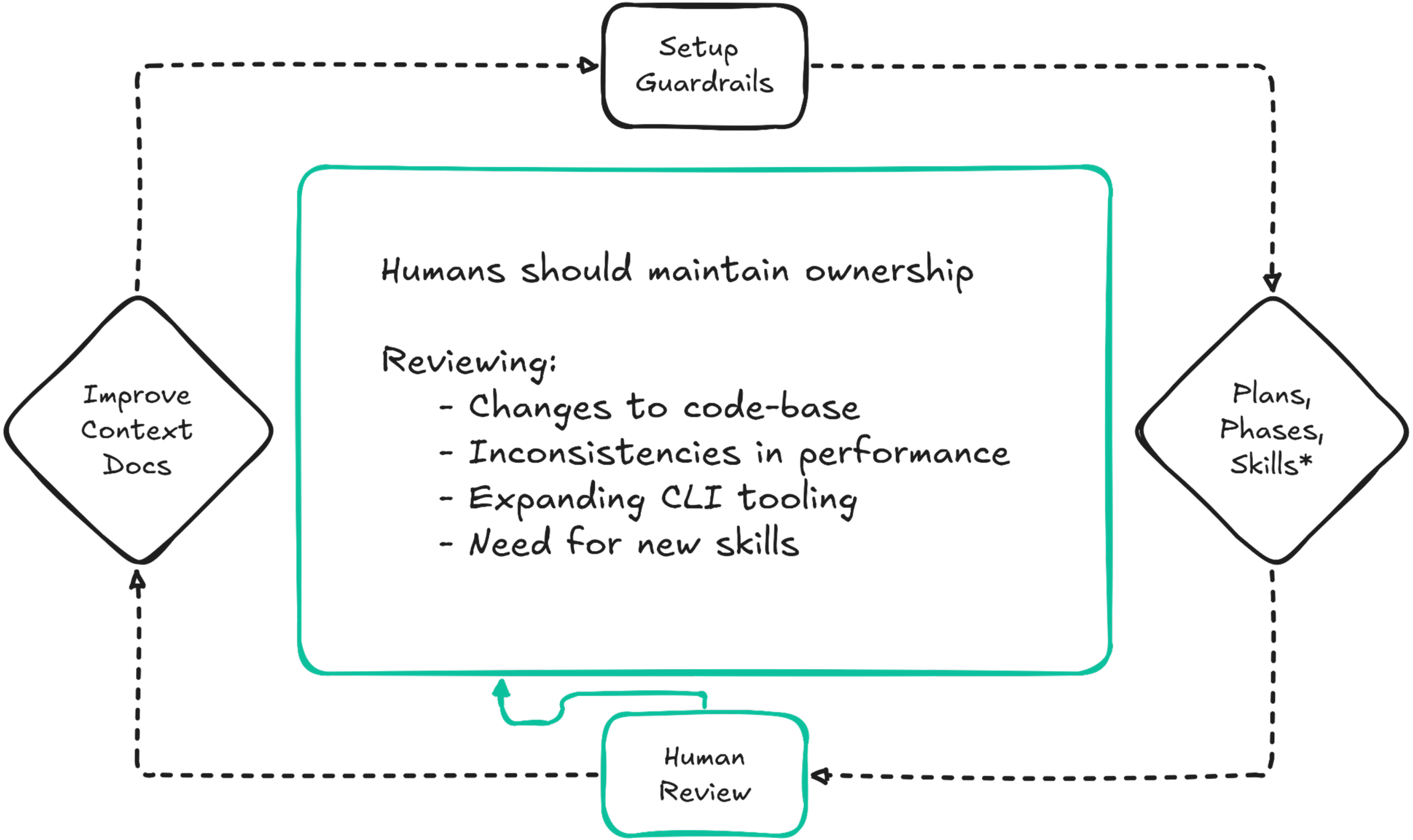
Plans, Phases, Skills

Human Review

Improve Context Docs

Keeping the footprint small
- Starting with a plan
- Working in phases

Skills for repeated tasks
- Users and Devs
- review-*
- update-*



Setup Guardrails

Plans, Phases, Skills*

Human Review

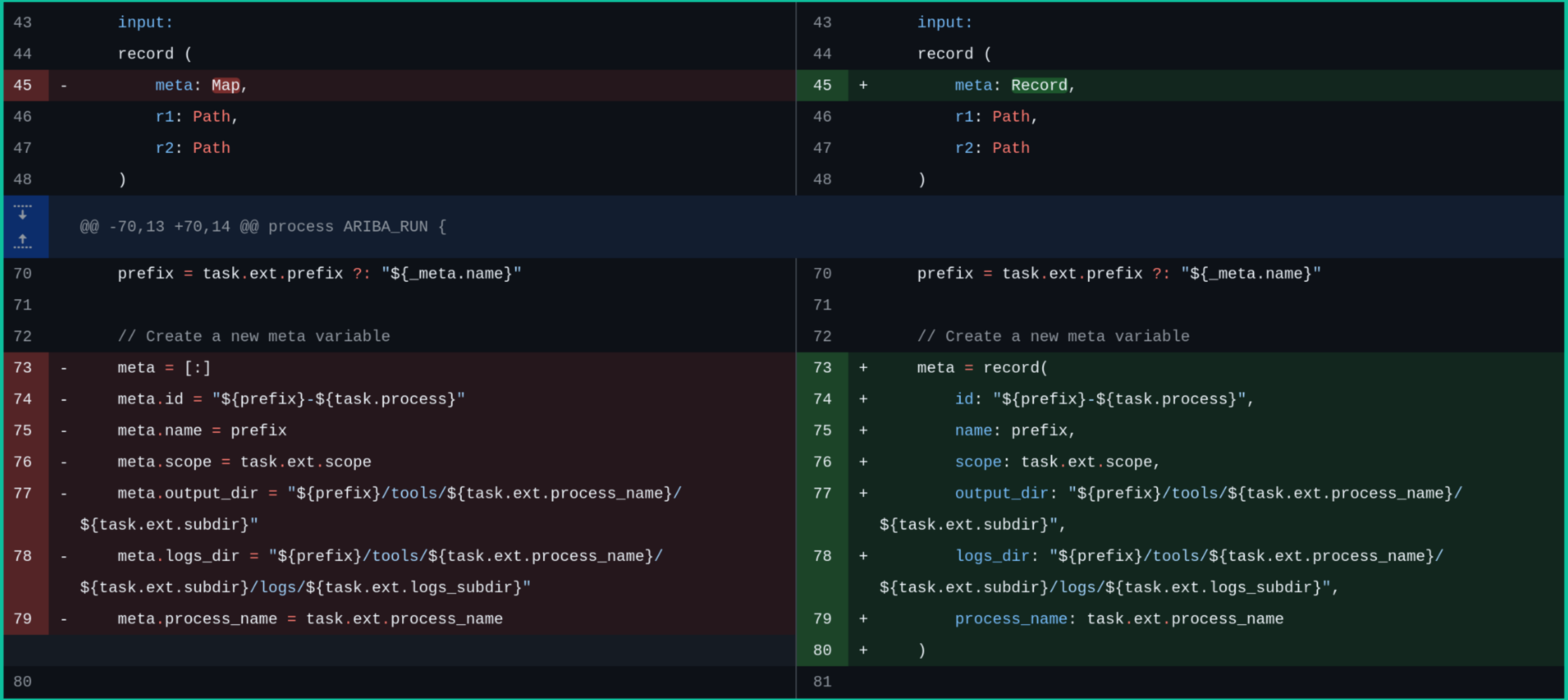
Improve Context Docs

Humans should maintain ownership

Reviewing:

- Changes to code-base
- Inconsistencies in performance
- Expanding CLI tooling
- Need for new skills

Feedback Loop in Action: meta Record Type Migration



```
    input:
    record (
-       meta: Map,
        r1: Path,
        r2: Path
    )
@@ -70,13 +70,14 @@ process ARIBA_RUN {
    prefix = task.ext.prefix ?: "${_meta.name}"

    // Create a new meta variable
-   meta = [:]
-   meta.id = "${prefix}-${task.process}"
-   meta.name = prefix
-   meta.scope = task.ext.scope
-   meta.output_dir = "${prefix}/tools/${task.ext.process_name}/${task.ext.subdir}"
-   meta.logs_dir = "${prefix}/tools/${task.ext.process_name}/${task.ext.subdir}/logs/${task.ext.logs_subdir}"
-   meta.process_name = task.ext.process_name

```

```
    input:
    record (
+       meta: Record,
        r1: Path,
        r2: Path
    )

    prefix = task.ext.prefix ?: "${_meta.name}"

    // Create a new meta variable
+   meta = record(
+       id: "${prefix}-${task.process}",
+       name: prefix,
+       scope: task.ext.scope,
+       output_dir: "${prefix}/tools/${task.ext.process_name}/${task.ext.subdir}",
+       logs_dir: "${prefix}/tools/${task.ext.process_name}/${task.ext.subdir}/logs/${task.ext.logs_subdir}",
+       process_name: task.ext.process_name
+   )

```

Feedback Loop in Action: meta Record Type Migration

- ***Improve Context Docs***
  - Update examples
- ***Setup Guardrails***
  - Lints, tests, doc validation
- ***Plans, Phases, Skills***
  - modules -> subworkflows -> workflows
  - review-docs, update-docs, run-tests
- ***Human Review***
  - Walk through changes
  - Make needed revisions to loop

```
// Create a new meta variable
meta = record(
    id: "${prefix}-${task.process}",
    name: prefix,
    scope: task.ext.scope,
    output_dir: "...",
    logs_dir: "...",
    process_name: task.ext.process_name
)
```

Much improved experience compared to previous 2025 experiences!

Bactopia v4 is now available! ✨🎉🕺

Ready for the future of Nextflow

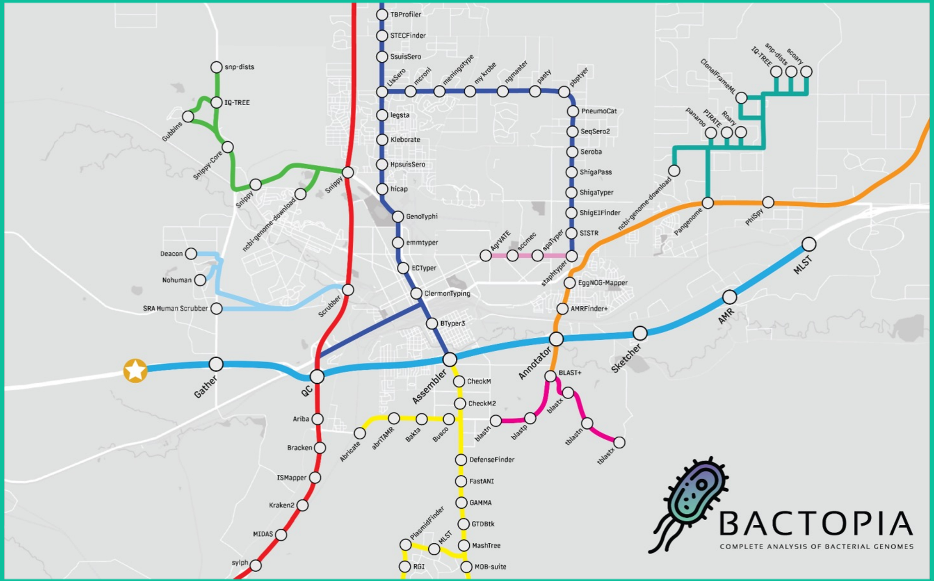
- Completely rewritten for 26.04+
  - Strict Syntax, Record Types
  - 2600+ files changed (*+138k+*, *-55k*)
- nf-test implemented for all steps
- Expanded ecosystem
  - Revamped bactopia-py
  - Created nf-bactopia Plugin
  - Docs built from GroovyDocs
- AI Ready Codebase
  - Machine readable indices
  - Skills for Users and Devs

BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

Acknowledgements

The many developers of open-source software and the users of Bactopia that regularly provide feedback.



WPHL

Taylor Fearing
Elena Groves
Maren Anderson
Chayse Rowley
Jim Mildenberger
Rob Christensen
Joseph Reed

Emory

Tim Read
EMERGENT Group

RUSH

Ahmed Babiker

Funders

WYOMING PUBLIC HEALTH LABORATORY
EMORY UNIVERSITY SCHOOL OF MEDICINE
cape Center for Applied Pathogen Epidemiology and Outbreak Control
CZ OPEN SCIENCE PROGRAM

nextflow SUMMIT
by seqera

Thank You!
Questions?

Ask me for a sticker or postcard!