

Building Open Source with LLMs

From purl-patrol to ecosyste.ms cli

Developing and Releasing OS Projects using Agentic LLM

From purl-patrol to ecosyste.ms cli

**Open Source
Teaching @ DC
Chameleon
Project Cars
Purl Patrol & 'The
Challenge'**

Origins

00

Open Source

was always there for me

Companies have been trying to figure out what it is that makes open source work. Brian B.

- **2000s PHP**
- **2005 CMS**
- **2010 Node**
- **2019 Depublish**
- **2025 ecosyste.ms**

Teaching @ MW

Level-Up the **DC** way

- was ~~pissed~~ not satisfied with the certification culture
- proposed find ways to practically empower people
- got my chance in 2023

Chameleon Projects

Practical Learning

- Collect people w.o project and start a 'larger than life' one
- Let everyone work in the aspired to role - not the actual one
- Create a space for deliberate practice - we assume to fail. We fail because its hard.

Project Cars

MLOPS

MLOPS

Complete AWS Stack including data ingestion, model training and inference via lambda

AUTOMOTIVE

Aspects of CAN bus data ingestion and other Automotive specialities

EDGE COMPUTING

Had some aspects of remote data collection in non optimal environments

IAC

No User Interface was harmed in the production of this App

Purl-Patrol

Automated OS License Compliance

Compliance is LAW

Several Laws get customers into situations where SBOMS are produced. This leads to further usage of the data in aspects like os licenses

Improve Available Tooling

Tooling about license compliance is often late to projects. These decisions are much less costly in the beginning

Creating Internal Standards

For regulatory reasons, it makes sense to help customer to compliant behaviour in terms of licenses. It is our responsibility to some extend.

Find new customer touchpoints

Need something else than AI or 'cloud' to sell? Customers need our help, but often dont know

purl-patrol

Results

- Working Software ([gitlab/azure](#))
- Intro to Bitkom OS Meetups
- OS Intelligence as Challenge

THE CHALLENGE:

Get OSInt data about dependencies **BEFORE** a build is starting into CI scripts

Why Them?

The CLI Project

Why LLM?

Ecosystems

01

Hello

ecosyste.ms

OS Int for Open Source
Knows the dependencies of your
dependencies, their licenses and CVES.
It has a api and tries to find a convenient way
to do this for any open source ecosystem.

ecosyste.ms

Why them?

- **Not part of security theater**
- **A true Open Source Infrastructure Project**
- **Keen to collab**

ecosyste.ms CLI

- Python CLI Interface to ecosyste.ms API
- Every subcommand one service
- Every Service a OpenApi Spec

The LLM Aspect:

Use an agentic engineering setup to create a CLI application in python as interface to OPENAPI documented backends.

Should be easy!

Should be easy!

(me all the

time)

Claude <3

Multiple iterations

Boring is good

Test driven

A codebase for humans

The dev process

02

Claude <3

- Small context: Code explains itself
- Claude got much better in building a cli from OpenAPI Specs

Multiple Iterations

Increment & Repeat

- 3 complete rebuilds
- not sure python was the perfect choice - given my skillset
- Surprise: had to learn ALL the services

Boring is good

I know what I want to do next summer!

- Dependency on the main service is hard. So why simulate smth. else
- Infrastructure does not need updates all 2 weeks

Test Driven

Red-Green-Refactor

- Python TDD is working quite well
- Uncovered: a myriad of internal cli errors, architecture problems, api bugs, scaling problems, broken concepts

A codebase for humans

The “no Schleiferl” approach

- This one goes into maintenance very soon. Only fixes.
- The uncovered challenges need addressing - not cli features

**ecosyste.ms asked to
move the CLI into their
project. Now released
as part of their software**

**No Vacuum
Cardinalities
Infrastructure**

Challenges

03

No Vacuum

- The ecosystem project was not built for an api
- Some quirks in API design from its Rails Backend
- 100% coupled to ecosyste.ms releases
- Which are coupled to google funding

Cardinalities

“Sea of Information”

- Lots of TMI Queries - “all wordpress CVE”
- Usage and load patterns are hard to predict with an app like this
- API not yet ready to take the load

Infrastructure

To make it really work

- API redesign for Idempotency
- Database Redesign for 'Internet Scale'
- Our own deployment with our own data

Deploy
Operate
Extend

Addendum

04

Deploy

- PYPI
- brew (??)
- other package managers?

Operate

Manage a new Open Source Project

- Considering asking for Non-Code Feature Requests instead of PRs
- Super stoked to see how much I can use the github mcp server

Achtung Alles Lookenskepers!
Das computermachine ist nicht
fuer gefingerpoken und
mittengrabben. Ist easy
schnappen der springenwerk,
blowenfusen und poppencorken
mit spitzenparken. Ist nicht
fuer gewerken bei das
leichtkopfen. Das rubbernecken
sichtseeren keepen das cotten-
pickenen hans in das pockets
muss; relaxen und watchen das
blinkenlichten.



Let's talk about MCP

**Let's talk about you
and me**

Extend: MCP

Hyper Hyper!

- Logical conclusion
- has no local state!
- But it is a proxy!

Extend: MCP

How to add the forbidden fruit!

- CLI usage patterns differ from MCP
- MCP being more resource intensive
- Rate limiting a mcp service limits dev speed.
- MCP likely to produce 'uncached queries'

Extend: MCP

- There are several engineering challenges to solve before we can mpc on a scale
- We need to test what agentic systems and their users ask

Global Options

- `--timeout` : Set the timeout in seconds for all HTTP requests (default: 20 seconds)
- `--format` : Set the output format (default: table). Available formats: table, json, tsv, jsonl

Output Formats

The CLI supports multiple output formats:

```
# Example: Set a 30-second timeout for all requests
ecosystems --timeout 30 repos topics
```



```
# Example: Get output in JSON format
ecosystems --format json repos topics
```

```
# Example: Get output in TSV format (tab-separated values)
ecosystems --format tsv repos topics
```

```
# Example: Get output in JSONL format (JSON Lines)
ecosystems --format jsonl repos topics
```

Contact:
Sebastian Schürmann

Questions?

Data

Core ecosyste.ms datasets

Packages

Metadata for 12.1m packages across 75 sources

[OpenAPI Spec](#)[Interactive Docs](#)[Service](#)

Repositories

Metadata for 287m repositories across 1,952 sources

[OpenAPI Spec](#)[Interactive Docs](#)[Service](#)

Advisories

Metadata for 24k security advisories across 12 languages

[OpenAPI Spec](#)[Interactive Docs](#)[Service](#)

Tools

Use ecosystems intelligence to get things done

Dependency Parser

Resolve the full dependency tree for a repository

[OpenAPI Spec](#)[Interactive Docs](#)[Service](#)

Dependency Resolver

Resolve the full dependency tree for a package

[OpenAPI Spec](#)[Interactive Docs](#)[Service](#)