

THE OPEN SOURCE CATALOG QUEST

Automating publiccode.yml · FAIR metadata · Code discovery

Robin Franken - (Swiss Data Science Center)
Sandino Scheidegger (Swiss Federal Chancellery)

THE KINGDOM'S BURDEN



The Scribe's Burden

Developers must hand-craft publiccode.yml metadata that already lives in the repository: A tedious, redundant ritual.



The Hidden Treasure

Without structured metadata, valuable public-sector code stays buried in the dungeon, difficult to discover or reuse.



The catalogue nightmare

Lots of software = Lots of choice, but not all repos are created equal! Think FAIR!

The realm cries out for heroes...

A STEPWISE QUEST



Rock of Alchemy

Foundation

- › Parse & map gimie output into publiccode.yml
- › First dataset: all Swiss gov repos provided



Elder-Wand of Automation

Enrichment

- › Add LLM layer on top of working pipeline
- › Parse free text fields (READMEs, descriptions)



Goblet of Judgement

Assessment

- › Score repos for FAIR compliance via gimie output
- › Community defines what makes a repo truly FAIR

Guild Laws & Legacy

 PRs to existing tools welcome, reviewed by guild masters

 Tests & CI must pass before merging

 AI tools permitted, but the knight controls the sword - do not submit code you cannot defend

 No dark magic. Hallucinations must be thwarted

The Legacy Lives On

gimie powers SDSC projects today. The SDSC team will continue to maintain & extend artifacts after the hackathon ends.



ANSWER THE CALL



Our Fellowship

- ◆ Python & Knowledge Graphs (gimie)
- ◆ RDF / Turtle & semantic web
- ◆ Open source catalog standards (Swiss, EU)
- ◆ SDSC infrastructure & deployment



Adventurers Sought

- ◆ YAML & Python warriors
- ◆ DevOps / CI pipeline experience
- ◆ Open source community veteran
- ◆ FAIR data & metadata enthusiast