The open-source & self-hosted **Data Lineage Platform**

Backed by **Y Combinator**

INVEST IN **GRAI (YC S22)**

Open source version control for metadata

LEAD INVESTOR

Max Keenan

I invested in Grai because Ian & team are solving a problem that I face: information inconsistency & data lineage errors plague companies of all sizes. It got so bad that when I was at ServiceNow, we had two separate teams present the same KPI and have vastly different numbers. Teams cannot make good decisions without good data. Grai's open source approach makes complete sense; it combines Ian's experience building and contributing to open source projects with a low-friction entry point into enterprises. They're a great team, solving a real problem and I'm excited to be along for their journey.

Invested $5,000 this round

grai.io   San Francisco California   [in]   Technology   Y Combinator   SaaS   Devtools   Ycs22

Highlights

1. Built Visions, a python OSS library with 13m+ downloads

2. Realized the need while working at Centene (F50 health insurer)

3. First line of code 2 months ago, already have pilot customers

4. $3.5b+ market opportunity

5. $519k raised from investors including a notable Business School prof. at Olin University

Our Team

Ian Eaves Founder & CEO

Ex-physicist turned ML engineer with experience building MLE teams everywhere from startups (Bellhops) to fortune 50 companies (Centene). An active open source contributor, he's written software with over 11M downloads (Visions)

Tony Edwards COO

Edward Louth CTO

Pitch

TL;DR Grai makes developers smarter by bringing metadata from across their stack into their development tools. We make testing data flows between applications painless.
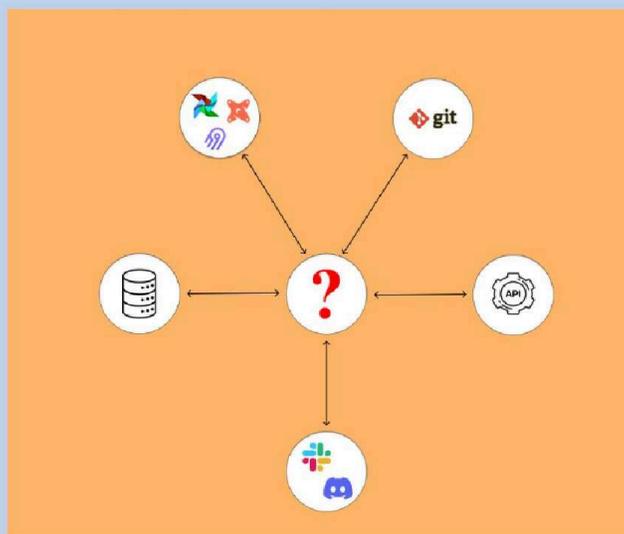
😭 Problem

Tracking (let alone testing) data once it leaves a production environment is challenging. Whether in a data engineers transformation pipeline, an ML model, or the CFO's metrics dashboard, that data *will* be used elsewhere. Without visibility on those use cases data changes remain risky and outage prone.


🚀 Solution

Grai is an open-source data management platform designed to help you better use your data.

- **Automated data lineage** - Pre-built connectors to keep metadata fresh.

- **Integrated with git** - Changing a column in your DB? Run data integration tests for all downstream users as part of your standard CI/CD process.

- **One-stop-shop** - A unified model of your entire data stack.

- **Your data belongs to you** - Grai is open-source & self-hosted with a cloud option coming soon.

Grai

## OSS version control for metadata

---
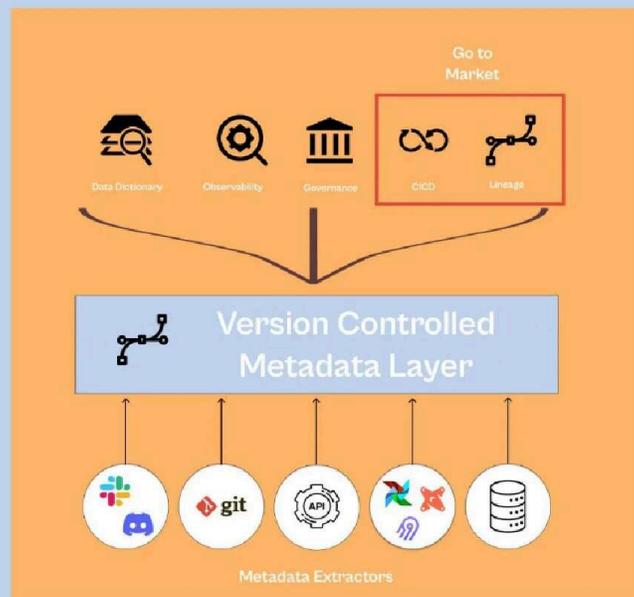
With Grai, customers understand how their data connects together
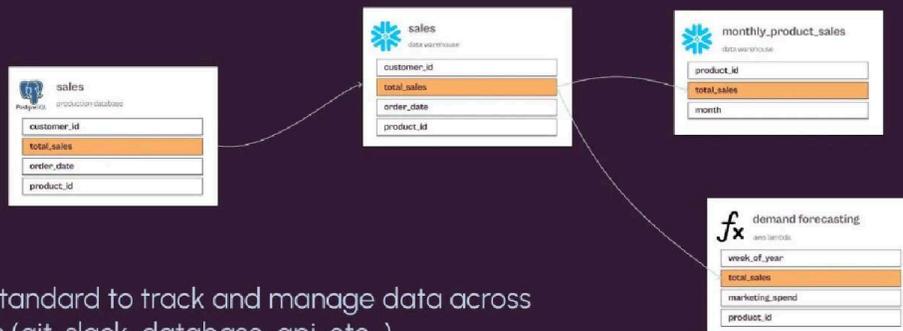
- A single standard to track and manage data across platforms (git, slack, database, api, etc...)

- Automatically works with customers existing tools

- Cross application data discovery and testing

Nothing to first customer in 2 months

Timeline

Jun	Jul	Aug
First commit	Limited open source launch	First customer signed

---
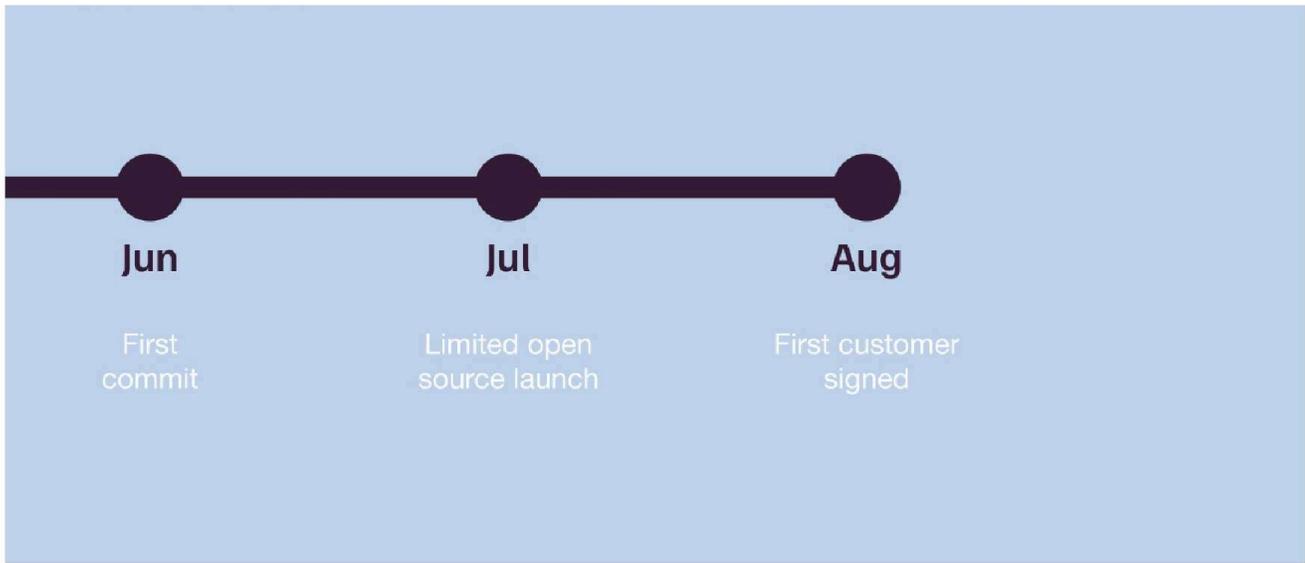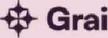
$3.5B
TAM

100k+ Companies with data warehouse x **$35k** Yearly ACV