**User:**

以下内容是准备做一个desci生态的信息库的一部分思路：

# 目的

* 为组织提供全球范围的desci项目、组织和行业的动态信息

# 任务

1. 做全球desci项目的信息档案库，每日更新；
2. 信息档案库作为原始数据，从中提炼和编辑有价值的信息，输出至国内各大主流媒体平台。

# 主要内容来源渠道

1. 主流生态：

* **molecule** **etica**
* **ResearchHub** **deScier** **DeSci Labs**
* **protocol Labs**

1） **molecule**生态的信息搜集：

* molecule labs平台中：desci项目发布、筹款、投资人和项目方谈判、代币化、研究成果发布各阶段信息；
* Bio Protocol 的‘BioDAO’发布及代币ICO、IDO情况，知名、潜力‘BioDAO’的代币市值、社区动向、产品进度及成果信息，
* Bio原生代币流通情况及组织资讯；
* 其他有价值信息

2） **etica**平台的信息搜集：

* 科研组织与etica合作动态信息：
* 已经建立的研究项目的成果发布、投票结果、代币奖励信息跟踪；
* etica原生代币跟踪；
* 其他有价值信息；

3） **ResearchHub**平台（“科学界的GitHub”）的信息搜集：

* 热门研究、收入和融资机会信息；
* 论文同行评审即可获得 RSC；
* 科研资助与资金支持信息：
* RH期刊阅读并发表同行评审的研究信息；
* 其他有价值信息；

4） **deScier**平台的信息搜集：

* 论文发布信息；
* 组织活动信息；
* 其他有价值信息；

5） DeSci Labs平台的信息搜集：

* 发表、审阅和策划研究成果信息；
* 利用 Web3 技术的科学出版解决方案信息；
* 利用web3的研究管理平台信息；

6） **protocol Labs**平台的信息搜集：

* web3，AI基础设施，行业动态信息跟踪；
* AGI、RWA 等前沿科技领信息

我想让你帮我从这个思路里面总结一下，信息档案库有哪些知识的维度，或者说如何建立信息档案库，选择什么样的产品去建立比较合适？请仔细作答，谢谢

**Kimi:**

一、信息档案库的“知识维度”  
把 DeSci 生态每天产生的大量碎片化信息沉淀为可检索、可分析、可二次创作的结构化知识，需要先把“维度”定好。下面给出可直接落地的 8 大一级维度、36 个二级维度，全部字段可用 JSON/CSV 统一存储。

| 一级维度 | 二级维度（字段） | 关键说明 |
| --- | --- | --- |
| 1. 基础档案 | 项目/组织 ID、中英文名称、官网、Logo、一句话定位、成立时间、主体法律结构 | 形成唯一标识 |
| 2. 赛道标签 | 研究主题（如长寿、脑科学）、技术栈（IP-NFT、DAO、ZK、AI Peer-review） | 方便后期做赛道热度或投资组合分析 |
| 3. 投融资 | 轮次、金额、代币/股权比例、领投方、跟投方、估值、代币经济模型文档 | 追踪资金走向 |
| 4. Token 动态 | 代币符号、合约地址、发行总量、流通量、价格、市值、24h 交易量、持币地址数 | 每日自动拉取链上数据 |
| 5. 治理与社区 | Snapshot 提案标题、投票结果、论坛热帖、Discord/Telegram 人数、Twitter 粉丝 | 评估 DAO 活跃度 |
| 6. 研究进展 | 阶段（概念→筹款→实验→论文→专利）、里程碑日期、关键成果 DOI、GitHub 代码库 | 与知识图谱打通 |
| 7. 出版与声誉 | 预印本链接、评审报告、被引次数、Altmetric 分值、开放数据徽章 | 用于科研影响力分析 |
| 8. 事件日历 | 会议、AMA、IDO、实验上线、论文见刊、治理投票截止日 | 支撑每日资讯快讯 |

二、如何建立信息档案库（工具选型+落地流程）  
根据“每日更新”和“二次内容输出”两大需求，推荐“ELK + API 爬虫 + 低代码 CMS”组合，兼顾自动化、可扩展、易二次编辑。

1. 数据层  
   • **Elasticsearch**（阿里云托管或自建）做全文检索与时间序列存储；  
   • **MongoDB** 存半结构化文档（JSON）；  
   • **IPFS/Filecoin** 做 PDF、图表、视频等大文件去重存储；
2. 采集层  
   • 通用爬虫：**n8n / Make.com / Apify** 无代码拉取官网、Medium、Mirror、Twitter、Discord；  
   • 链上数据：**The Graph / Dune API / Covalent** 直接拿 Token、治理投票数据；  
   • RSS+Webhook：ResearchHub、DeSci Labs、DeScier 有公开 RSS 时，用 **Logstash RSS input** 直接进库；  
   • 定制脚本：对 Molecule、BioDAO 这类无公开 API 的站点，用 Puppeteer 或 Playwright 定时爬取并写回库。
3. 处理层  
   • **Logstash / NiFi** 做字段清洗、去重、打标签；  
   • **LLM Pipeline**（OpenAI + LangChain）自动提取“一句话定位”、研究主题、投融资金额；  
   • 版本控制：每条记录加 updated\_at，并保留 7 天 diff，方便回滚。
4. 展示/二次创作层  
   • **Strapi / Directus** 头 CMS，编辑可低代码增改记录；  
   • **Grafana + Kibana** 做实时仪表盘：赛道热度、Token 市值、DAO 活跃度一键可视；  
   • **Notion API** 同步每日“精选 5 条”卡片，供国内小编二次编辑后分发到公众号、微博、36kr、财新。
5. 自动化流程  
   • 每日 0:00 UTC 启动爬虫 → 1:00 完成清洗入库 → 2:00 LLM 生成中文摘要 → 3:00 推送到 Notion 草稿箱；  
   • 重大事件（融资>1M 美元、Token 上市、治理提案通过）触发企业微信/Slack 实时告警。

三、小结  
• “维度”解决“存什么”；  
• “ELK + 爬虫 + CMS”解决“怎么存、怎么每日更新”；  
• 低代码链路让编辑团队无需工程师也能每天把最新 DeSci 动态快速分发到国内主流平台。