Jonathan Starr

Burlington, VT · jon@scios.tech · jstarr.science · LinkedIn

SUMMARY

Thirteen years of open infrastructure for open science and open source, as builder, author, funder, convener, and teacher. Co-Director of SciOS (501c3), Instantiating Steward of IOSP, and Founding Executive Director of the Open Source Endowment (501c3), the first endowment fund exclusively for open source software. I build production systems for open science, including PRSM (a knowledge graph of 78M+ relationships across scientific work), GREP (translating the existing corpus of scientific literature into frontier systems such as modular research objects, beginning with software mentions extraction from pdfs), RRGI (a publishing system for signed scientific claims hosted on resilient distributed infrastructure), the IOSP Node Network (researcher-owned IPFS data-preservation infrastructure launching at IOSP 2026), Lettuce Compute (open infrastructure for volunteer computing), as well as the multi-agent development system that ships all of it. I design the frameworks beneath them, including the core-satellite funding model and *Resilient Data Futures*. I fund the commons, through the Open Source Endowment ($750K+ committed from 108 donors) and grant review for the U.S. DOE and NSF. And I convene and teach the communities building the rest, through IOSP and PICoding (training researchers to build multi-agent AI workflows), and I contribute to community-stewarded standards like MIRA.

BUILDING

PRSM *(builder, beta)* — prsm.network Living knowledge graph of research and digital infrastructure. 21.8M entities and 78M relationships drawn from OpenAlex, GitHub, GitLab, ecosyste.ms, 15 package ecosystems, and 77M software mentions across ~24M papers; 15 premade analysis algorithms, custom impact scoring, and bulk import up to 10,000 DOIs or repository URLs per run. Institutional discovery starts from a university's name and web domain, hunts ten independent signals, and returns confidence-scored, human-reviewed repositories; the first full run surfaced 17,538 repositories across all 35 CURIOSS member institutions. PRSM also maps the MCP ecosystem, deduplicating seven registries into a graph of 58,000+ servers and 74,000 tools, so anyone can look up what AI tooling can reach a given dataset or library. Live demo portfolios include NumFOCUS, CZI EOSS, Bioconductor, the Fundamental AI labs, and the CURIOSS member universities.

Lettuce Compute *(creator, open beta)* — github.com/jring-o/lettuce-compute · beta: scios.tech/lettuce-beta Open infrastructure for volunteer computing, unlocking 6000+ ExaFLOPS of processing power. A researcher deploys their own head, defines computations on it, and volunteers from around the world complete the computational work for free. Results are cross-checked between independent volunteers before they count, and every credit decision is published as a signed attestation a third party can verify without trusting the head that issued it. In open beta since July 2026, running a live volunteer fleet across GREP and a BeyBlade simulator.

SciOS Compute *(builder, live)* — compute.scios.tech Hosted Lettuce Compute with the setup done for you. An AI-assisted onboarding pipeline takes a researcher from research question to distributable code running on volunteer machines. GREP runs on it today.

GREP — The Great Research Extraction Project *(builder, live in beta)* — compute.scios.tech Translation of the existing corpus of scientific literature into frontier systems, such as modular research objects. Its first campaign extracts every software mention in the literature, and is the first real scientific computation running on Lettuce Compute. Three independently fine-tuned DeBERTa-v3-large models, distilled from a 44 GB teacher ensemble analyze pdfs; a mention enters the record only when at least two of the three find the same span, and every work unit is run by three separate volunteers before its result is accepted. The crowd scores F1 0.833 on SoMeSci dev, 0.828 on SoMeSci test, 0.865 on Softcite, and 0.894 on SoFAIR, with a 2.2% hallucination rate. The first public corpus, 2,520 papers across six domains, is loaded and running as of August 1st 2026.

RRGI — Resilient Research Graph Infrastructure *(architect and builder, live)* — graph.scios.tech A live publishing system for scientific claims built on decentralized identifiers, the AT Protocol, and IPFS. Researchers sign in with their own cryptographic identity and publish questions, claims, evidence, and typed relations to their own data servers; 24 custom lexicons, signature-verified ingest, and an enforced evidential grammar maintain the graph, and every record is archived to two independent IPFS nodes so any claim can be re-proved from cold storage with the origin server offline. Anyone can endorse or dispute any record, including machine-written links, through five revocable signed stances, and a record's standing is tallied at read time from those stances rather than granted by an editorial board. The public test-graph carries 5,897 archived records, including a signed six-paper corpus that passed an independent audit with zero illegal edges.

IOSP Node Network *(builder, reference deployment live)* — iosp.science/datanetwork Researcher-owned data-preservation network being stood up for IOSP 2026, funded by the IPFS Foundation's Implementations Fund. At the workshop, twenty researchers stand up their own IPFS nodes, half on grant-funded take-home Raspberry Pi kits, load them with at-risk datasets from their own disciplines, and leave with a published build runbook and agent skills so their labs can add nodes without us. The reference deployment has run continuously since July 30, 2026, as two parallel private IPFS clusters with no cloud and no DNS anywhere; three tiers of autonomous monitoring prove every node still serves its data, from 60-second heartbeats to half-hourly cryptographic proof-of-serving probes, and a public dashboard shows the network live.

Citey *(builder, live)* — citey.scios.tech Software citation toolkit. Chrome extension that auto-detects software in highlighted text and emits BibTeX `@software{}` entries; CFF and CodeMeta generators. Designed and built in one day with Agent Teams at the 2026 Software Citation Workshop (supported by NASA under award 24-TWSC24-0068).

Agent Teams and the Software Factory *(builder)* Agent Teams is a process-driven multi-agent orchestration system built on Claude Code and the Anthropic Agent SDK. A thin Lead dispatches ~17 specialized subagents through structured agent-to-agent handoffs; handoff files form an immutable audit trail, SQLite persists pipeline state, bash hooks enforce quality gates, and a custom MCP step-server enforces step ordering within each agent. Every line of code in GREP, Lettuce Compute, SciOS Compute, and PRSM ships through this pipeline. Its successor, the Software Factory, is in development. There, ten specialized agents work five lanes (bug, feature, chore, hotfix, new project); deterministic code rather than agents owns every quality gate; and bug and feature work is attempted three times over in isolated git worktrees before a judge agent picks a winner.

WRITING

The Core-Satellite Model *(author)* — scios.tech/thoughts A funding interface for open source and open science. A core of domain experts stewards a field's canonical artifacts and offers funders one accountable relationship; money moves under stated allocation rules to the satellite projects orbiting it; a minimal shared schema makes cores comparable, so legitimacy is computed by observers rather than granted centrally. Published with a hand-authored discourse graph (31 claims, 49 evidence nodes, 180 sources). A working session at UN Open Source Week in New York (June 2026) brought two dozen funders, fiscal sponsors, and infrastructure stewards into the model; pilot cores are forming in modular science and research infrastructure, with the first reporting results at IOSP 2026.

Resilient Data Futures *(author)* — rdf.scios.tech A case for treating research data loss as an architectural problem. The whitepaper sets out a Tier 0→3 resilience framework and estimates the latent liability of irreproducible data at roughly $1.1B per year for a representative R1 institution. It is published as a MIRA graph of 349 connected research objects, with the full paper and composed narrative views at rdf.scios.tech/narratives. A monthly working group of research software engineers, formed at US-RSE 2025, carries the program forward.

CONVENING & TEACHING

IOSP — Institute of Open Science Practices *(instantiating steward)* — iosp.science Annual production-driven convening on open scientific infrastructure; 75% of the program is hands-on building, 15% is the tooling showcase, and 10% is talks. The 2025 inaugural event in Denver drew 425 registrations for 80 curated seats from 5 countries; 95% of participants would attend again, and work begun there seeded MIRA, CAIROS, and PRSM. IOSP 2026 runs October 12–15 in Leiden, hosted by the University of Leiden and GO FAIR, with 100 participants building the open science stack across four tracks (Store, Share, Assess, Fund).

PICoding *(creator and lead instructor)* — scios.tech/picoding Four-week intensive teaching researchers to build production AI-agent workflows from their own data and processes, covering agent context, reusable skills, delegation, and orchestrated multi-agent harnesses. Cohort 1 (April 2026) was 4× oversubscribed. Cohort 2 starts in August 2026.

MIRA — Modular Interoperable Research Attribution *(contributor)* — mira.science An open schema for modular research attribution, stewarded by a community of 20+ researchers, tool-builders, and publishing professionals. 100+ researchers across 10+ laboratories publish through it in a typical week, across five independent tools and 6,000+ records. SciOS helps advance the schema; lead-organizer for the five-day design workshop in Ireland (June 2026, 22 invited participants, funded by The Navigation Fund) that refined the schema and seeded tool implementations.

EXPERIENCE

Co-Director, SciOS (501c3) — *Dec 2023 – present* · scios.tech With co-founder Ellie DeSota, built the organization from zero. Shipped PRSM, GREP, RRGI, Citey, SciOS Compute, and PICoding. Paid engagements include the Modular Science movement and the Quantum Biology Institute (FAIR publishing pipelines).

Founding Executive Director, Open Source Endowment (501c3) — *Feb 2024 – present* · endowment.dev The first endowment fund exclusively for open source software. Donations grow an invested principal and only investment returns fund grants, making the fund evergreen. Built the organization from zero, authoring the Form 1023, and governance charter. **$750K+ committed from 108 donors**, 65 of them governing Members ($1K+/yr).The first grant round, in 2026, will direct ~$2.5K micro-grants to independent open source projects under a data-driven Value-Risk framework.

Instantiating Steward, IOSP — Institute of Open Science Practices — *2024 – present* · iosp.science Co-founded and steward the institute; convene the annual gathering (Denver 2025; Leiden 2026, hosted with the University of Leiden and GO FAIR).

Program Manager, Open Source Science Initiative (OSSci) — NumFOCUS — *Jul 2023 – Jul 2025* Grew OSSci to 600+ individuals across 19 countries. Led the Map of Open Source Science across 120+ projects and 40+ organizations.

Managing Consultant, DeSci Labs — *Oct 2022 – present* Advise leadership on organizational structure, strategy, and technical roadmap for decentralized research-object infrastructure (dPIDs).

Founder & President, The Science Commons Initiative (501c3) — *Jan 2022 – Apr 2024* Citizen-science 501(c)(3); 200+ students contributed distributed compute to active research.

President, LBRY Foundation (501c3) — *Jan 2020 – Dec 2022* Designed and led three quadratic funding rounds across a ~$1M grant purse.

Core Contributor, Gridcoin — *2013 – present* Authored the foundational whitepaper for a multi-role incentive protocol (computational, research, funder, storage) coordinating participants with differing incentives, capabilities, and operational contexts. The mechanism has run on 5,000+ nodes for 12+ years, with ongoing design work on temporal-adjustment functions for delayed-impact funding rewards.

GRANTMAKING & GRANT REVIEW

- **U.S. DOE ORISE — Reviewer** *(2025)* — CORSA, PESO, S4PST, STEP
- **NSF POSE — Reviewer** *(2024)*
- **Digital Infrastructure Insights — Reviewer** *(2023–2024)*

Foundation-side, I have designed and operated grant programs at the Open Source Endowment ($750K+, Value-Risk framework, first round 2026) and the LBRY Foundation (~$1M across three quadratic funding rounds, 2020–2022).

SELECTED TALKS

- **U.S. DOE Open-Source Software Summit 2026** — keynote
- **OSCON 2026** — *Funding Open Source as a Public Good*
- **PIDFest 2026** — *PRSM: What Persistent Identifiers Make Possible at Scale*
- **IRSC 2026** — *From Illegible to Investable: Tooling the Digital Commons for the Coming Decades*
- **All Things Open 2026** — *The Future of Science Depends on Open Source Engineers*

- **DWeb Camp 2026** — *Persistent, Atomic, Compositional Science: Foraging the DWeb Substrate for the Scientific Record*
- **UW–Madison iSchool 2026** — guest lecture on PRSM
- **OFA Symposium 2025** — *One Year Later: Building the Open Source Endowment*
- **All Things Open 2025** — *Why Open Source Needs an Endowment*
- **FOSDEM 2025** — *Building Bridges Between Researchers, Technologists, and Infrastructure* (Open Research Track); *Bridging the Gap: Regional OSPO Networks* (Community Track)
- **Wolfram ChainScience 2023** — *The Future of Science* (keynote)

EDUCATION

University of Rochester — BA in English, BA in Film Studies (2007–2011)