IPEDSBrain

Agentic Research Infrastructure for Institutional Intelligence

John Dimatos — February 2026

The Problem

Institutional Research offices face the same bottlenecks at every university:

Data assembly is manual and repetitive

IPEDS alone is dozens of survey components across multiple years. Every new question means re-joining admissions, outcomes, enrollment, financial aid, and completions data from scratch.

Reproducibility is fragile

When the Provost asks "how did you get that number?", the answer is a chain of Excel operations that are difficult to retrace. When an accreditor asks for a different peer group, the work starts over.

Reports are a separate bottleneck

The Problem (continued)

Methodology is inconsistent

Different analysts define peer groups differently, apply different filters, use different IPEDS vintage years — without documenting those choices.

Results vary in ways that are hard to detect.

The environment is getting harder

- NCES has been reduced to **3 staff members**
- New IPEDS reporting requirements are expanding for 2025-2026
- EDUCAUSE named "The Data-Empowered Institution" the **#1 IT issue for 2025**

The tools your IR office depends on may not be maintained. The reporting demands are growing.

What IPEDSBrain Does

A command-driven research tool your IR team operates through natural language:

```
ipeds:profile Arizona State University
```

→ Comprehensive institutional portrait

```
ipeds:benchmark --peers big-ten
```

→ Peer comparison across all metrics

```
ipeds:trends graduation_rate --portfolio our-aspirants
```

→ Multi-year trend analysis

```
ipeds:equity
```

Every Analysis Produces Three Things

1. A narrative report

Every claim cites its specific data source — IPEDS table, field, year, institution.

2. A reproducible script

Standalone Python + pandas. Recreates every number without the AI framework. A colleague runs `python analysis.py` and verifies the results.

3. A data bundle

The exact data slices used, packaged with methodology documentation. Hand the zip to an auditor or accreditation reviewer — they verify independently.

> No AI, no proprietary software, no API keys required to check the work.

The Reproducibility Bundle

```
analysis-2026-03-15.zip
├── README.md                 # How to reproduce this analysis
├── report.md                 # Narrative report with inline citations
├── analysis.py               # Standalone Python script (pandas only)
├── data/                     # Exact data slices used
├── methodology.md            # Analytical choices documented
└── citations.json            # Every claim traced to source data
```

Level 1 — Check the math: Run the script, verify every number.

Level 2 — Challenge the assumptions: Read the methodology doc, understand why those filters and definitions were chosen.

Architecture

Three layers, each with a clear responsibility:

Layer	What It Does	Key Property
Data	Fetches and caches IPEDS, Scorecard, and federal sources	Deterministic, auditable
Analysis	Specialized agents produce findings with mandatory citations	Every claim validated against actual data
Interface	Claude Code skills expose capabilities as simple commands	No Python knowledge required

Data sources: IPEDS (bundled, multi-year), College Scorecard (API), extensible to NSLDS and Common Data Set.

What It's Not

It's not a dashboard

Dashboards show data. IPEDSBrain answers questions — and shows its work.

It's not a data warehouse

No database to maintain, no ETL pipelines to monitor. The data layer is bundled and portable. It runs on a laptop.

It's not a black box

Every analysis generates a standalone script that reproduces every number without the AI framework. The methodology document explains every analytical choice. The citation file traces every claim to a specific field in a specific federal dataset.

It compounds over time

Who Uses It

Role	How They Use It
IR Analysts	Run analyses daily — profiles, benchmarks, trend reports
IR Director	Define peer groups, set methodology standards, review outputs
Provost	Receive institutional profiles and peer benchmarks for strategic planning
Deans	Program-level analyses — growth, demographic shifts vs. peers
Enrollment Management	Admissions yield, competitor analysis
Accreditation	Pull equity audits, outcomes data, methodology

Delivery: Phase 1

Core Engine + First Report (3-4 weeks)

Deliverables:

- IPEDSBrain installed and configured
- Bundled IPEDS data (2019-2023, all survey components)
- College Scorecard integration (earnings, debt, completion by income)
- `ipeds:profile` and `ipeds:benchmark` operational
- 3-5 custom peer portfolios built with your IR team
- **First report:** institutional profile with peer benchmarking, ready for Provost review

Working sessions: 2-3 per week with the IR team — defining peers, validating methodology, reviewing outputs.

Delivery: Phase 2

Trends, Equity + Dashboards (2-3 weeks)

Deliverables:

- Multi-year trend analysis with anomaly detection
- Equity audit — outcome gaps across demographics
- Interactive dashboard generator (standalone web dashboards)
- **Second report:** trend analysis and equity audit, plus a standing dashboard

Why this matters for accreditation

Most regional accreditors require evidence of outcomes tracking over time and demonstrated commitment to equity in student outcomes.

This phase produces exactly that evidence, with full reproducibility.

Delivery: Phase 3

Output Templates + Training (2-3 weeks)

Objective: Turn IPEDSBrain from "something the consultant runs" to "something your IR office produces routinely."

Deliverables:

- Report templates customized to institutional branding
- Slide deck generator for board presentations, accreditation self-study, internal planning
- Portfolio management training
- IR team proficiency in all commands
- **Third deliverable:** board-ready presentation generated by your IR team — not the consultant

Delivery: Phase 4

Handoff + Retainer Transition (1-2 weeks)

Deliverables:

Document	Audience	Purpose
User Guide	IR team	Command reference, workflows, portfolio management
Methodology Reference	Analysts, auditors, accreditors	Data sources, definitions, statistical methods, limitations
Architecture Document	IT staff	System design, data flow, extension points, security
Runbook	IR team + IT	Data refresh, troubleshooting,

Ongoing Support

Quarterly

- IPEDS data refresh when new vintage releases
- Methodology review for IPEDS survey changes

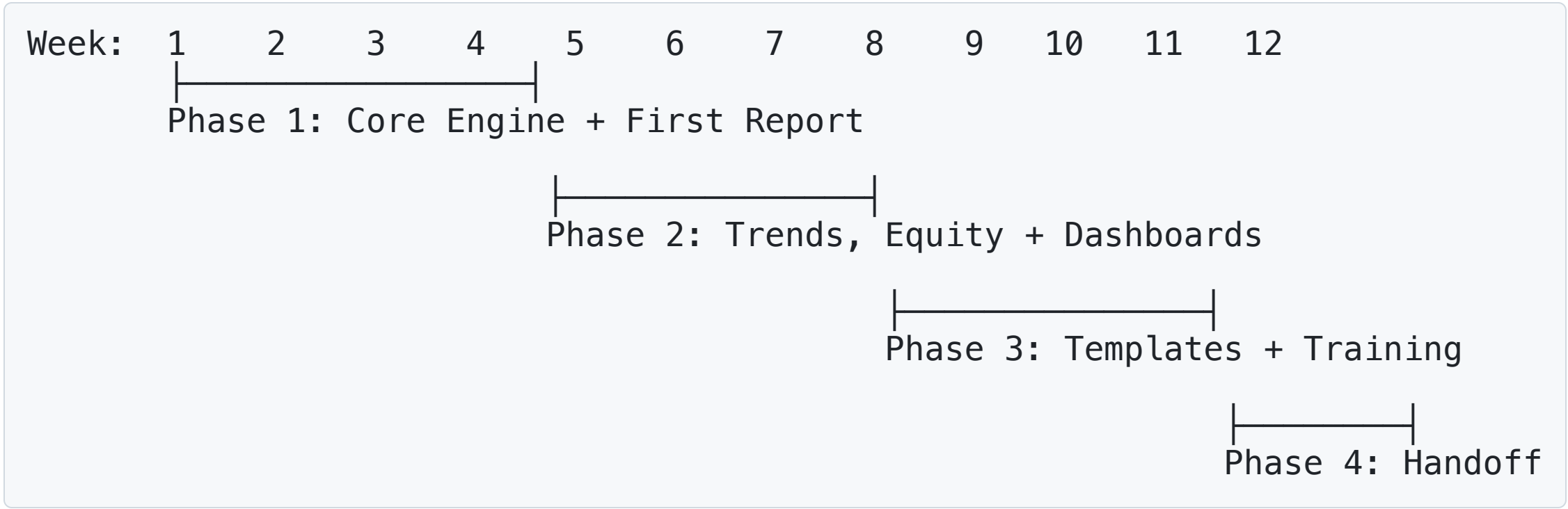
On demand

- New analyses for emerging institutional questions
- Additional peer portfolios as strategic context shifts
- New data source integration (NSLDS, Common Data Set, state data)
- Accreditation self-study support

Annual

- System review and dependency updates

Timeline



Total: 8-12 weeks. Each phase produces immediately usable deliverables.

Embedded partnership during build → ongoing support retainer after handoff.

Technology

Component	Technology
Runtime	Claude Code (Anthropic)
Language	Python 3.10+, pandas, pydantic
Data sources	IPEDS (bundled), College Scorecard (API), extensible
Visualizations	Observable Framework (interactive web dashboards)
Presentations	Marp (markdown slide decks)
Reports	Markdown + PDF via pandoc
Infrastructure	A laptop with Python and Node.js

No servers. No databases. No cloud services. No vendor lock-in.

Reproducible output bundles use only pandas and standard Python — your analyses

Summary

An AI-assisted research tool that answers questions, shows its work, and lets anyone verify the results.

8-12 weeks to deliver. Runs on a laptop. Every analysis is reproducible.

Pricing, dates, and terms to be discussed.