ESCALATE

NSF CyberTraining · Award OAC-2519416 · (Jul 2025 – Jun 2028)

Engaging Scholars in Cybersecurity Analysis: A Laboratory for Teaching and Education

Hands on cybersecurity training on NSF cyberinfrastructure: a shared catalog of AI-ready assignments built on real Internet measurement data.

kc claffy [1], Ricky Mok [1], Bradley Huffaker [1], Alexander Marder [2], Rocky Chang [3]
[1] CAIDA / SDSC / UC San Diego · [2] Johns Hopkins University · [3] Calvin University

caida.org/projects/escalate

Motivation

- Most cybersecurity training stops at theory and protocols; few programs let students tell baseline Internet behavior from real attacks in measurement data.
- Real datasets may be terabyte-scale: the UCSD Network Telescope records over **1 TB of packets per day**; 20 years of global routing snapshots fill **78 TB**. Too large, and often too sensitive, for laptops.
- Educators often lack cyberinfrastructure (CI) skills, so NSF-funded compute goes unused in courses.
- No shared catalog of CI-ready teaching material.

Goals

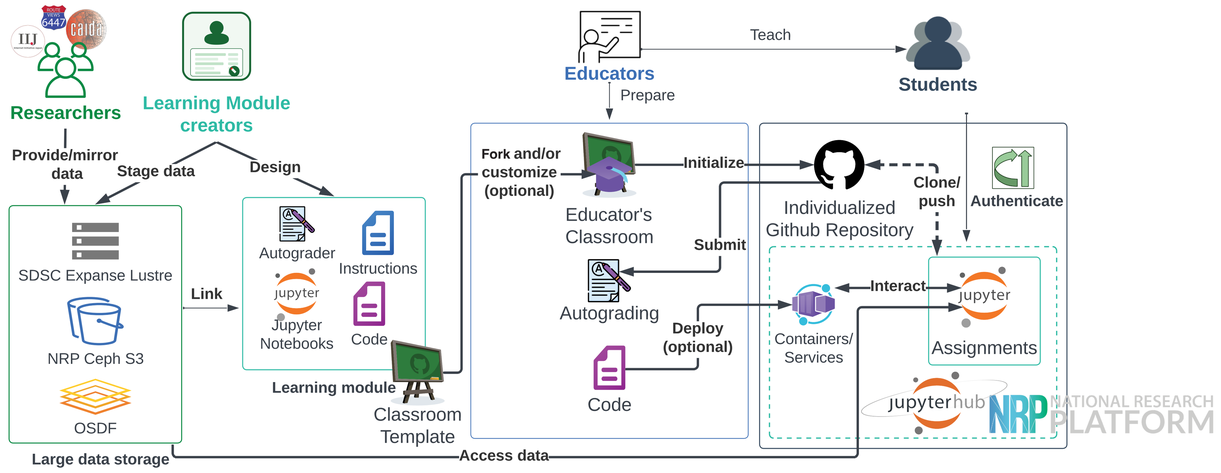
Three pillars, built on CAIDA's Internet Data Science for Cybersecurity course:

1. **Data-driven assignments** on real, Internet-scale datasets.
2. **Infrastructure to run and grade them** at class scale: NRP JupyterHub, SDSC Expanse, autograding.
3. **A shared catalog (C²Hub)** where educators find, adapt, and contribute modules.

Partners: UC San Diego, Johns Hopkins, Calvin University
. Seven more institutions have committed to adopt the assignments.

Approach: C²Hub on NSF cyberinfrastructure

(1) Researchers stage data on NRP, Expanse, and OSDF. (2) Module creators link data to notebooks, code, autograders. (3) Students clone assignment into Juypter-Hub on NRP and push results back.



- Datasets live on NSF cyberinfrastructure (NRP Ceph S3, SDSC Expanse, OSDF).
- Assignments run on NRP's hosted JupyterHub; students need no local setup.

What an assignment is

Two GitHub repositories under the CAIDA organization: public `nids-<topic>` (Introduction, Datasets, tool guides, Tasks, student notebook) and private `nids-<topic>-key` (worked answer key, data-loading scripts, container manifests).

Planning, content, and style guides let outside contributors build assignments.

Educator workflow

1. Find an assignment in the catalog
2. Create a GitHub Classroom assignment
3. Edit the template if needed
4. Create an NRP namespace and add students
5. Grade with the key

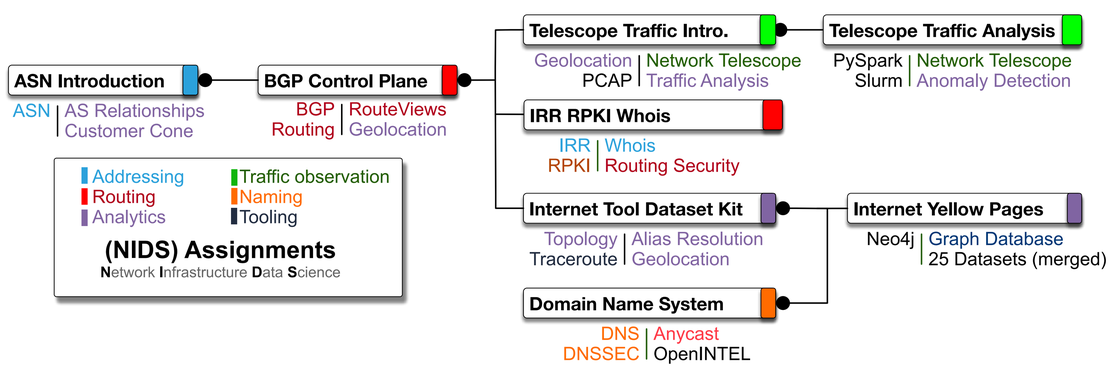
Student workflow

1. Clone assignment from GitHub
2. Join the NRP namespace with campus credentials
3. Start JupyterHub
4. Work from the template notebook
5. Push to submit

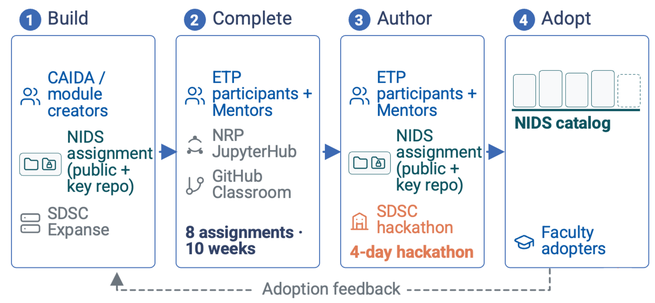
Assignment catalog

Eight public assignments, each with a private answer key

CODE	ASSIGNMENT	DATASETS AND TOOLS	RUNS ON
ASN	**ASN Introduction**	AS-to-Organization, AS Relationships, Customer Cone	NRP
BGP	**BGP Control Plane**	RouteViews RIB snapshots, prefix-to-AS, multi-origin AS (MOAS) prefixes	NRP
IRR	**Registries: WHOIS, IRR & RPKI**	RIR WHOIS/RDAP, IRR route objects, RPKI ROAs vs. observed BGP origins	NRP
ITDK	**Internet Topology Data Kit**	Router-level topology, alias resolution, geolocation (Postgres)	NRP
DNS	**DNS Ecosystem**	OpenINTEL active DNS, LACeS anycast census, DNSSEC adoption	NRP
TELESCOPE	**Telescope Traffic Introduction**	UCSD Network Telescope packet captures → flows, geolocation, origin AS	NRP
UCSDNT	**Network Telescope Analysis**	FlowTuple time series, PySpark → Slurm job array, country-level blackedout detetion	SDSC Expanse
IYP	**Internet Yellow Pages**	IYP Neo4j knowledge graph, Cypher queries, RPKI vs. BGP	NRP



Community: ESCALATE Training Program 2026



CAIDA built NIDS assignments, ETP participants completed them; participants authored new assignments at hackathon; new assignments will join the NIDS catalog for faculty to adopt.

Pilot cohort, Jun 15 – Aug 21, 2026 [4]: nine remote weeks plus a four-day hackathon at SDSC (Aug 18–21). Undergraduates, graduate students, postdocs, and research collaborators; mentors at 4:1 or better.

The eight assignments were the syllabus, one per week, building toward a capstone question: can we detect and attribute a country's connectivity degradation from routing, topology, and tralc data?

Week 9: build your own. Each participant planned a new assignment with the contributor guide, then drafted it at the hackathon with a mentor.

Hackathon: 30 attendees, including 10 mentors and 18 students. Six project teams (tralc, DNS and BGP security, AI Ops, active measurement, BGP and router topology); guests from Internet2 and RouteViews.

Incentive to produce: stipends to first authors of assignments that a faculty member uses next year.

Outreach: paper, *From Module Consumers to Module Producers*, ACM SIGCOMM Workshop on Networking Education for the AI Generation [2]

Contact training-program-info@caida.org · caida.org/projects/escalate/ · github.com/CAIDA (`nids-*` repositories)

This material is based upon work supported by the U.S. National Science Foundation under OAC-2519416. Any opinions or bndings do not necessarily recect the views of the NSF.

[1] ESCALATE proposal, NSF CyberTraining award OAC-2519416, 2025. caida.org/funding/cybertraining-escalate/

[2] kc Claffy, Ricky K. P. Mok, Alexander Marder, Bradley Huffaker, and Rocky K. C. Chang, "From Module Consumers to Module Producers," Presented at the ACM SIGCOMM workshop on Networking Education for the AI Generation.

[3] NIDS assignments and how-to guides. caida.org/projects/nids/

[4] ESCALATE Training Program (ETP). caida.org/workshops/training/2606/