

# Project Status Report

**Reporting period: 08/01/2022 - 03/31/2023**

**Project title:**

**Mid-Scale RI-1 Design Project (M1:DP):  
Designing a Global Measurement Infrastructure to Improve Internet Security  
(GMI3S)  
[OAC-2131987](#)**

**Principal Investigator:** kc Claffy, Bradley Huffaker (UCSD), David Clark (MIT)

**Project Manager:** Elena Yulaeva

**Lead Institution:** CAIDA, UCSD

**Other Institutions:** NSRC (U Oregon), MIT

**Cognizant PO:**

## Table of Contents

1.	Summary of project status	1
2.	Near-Term Milestones	1
3.	Technical progress highlights	5
4.	List of Acronyms	7

## 1. Summary of project status

A brief summary of the project's overall status on technical progress, cost and schedule performance.

Award Duration	Start date: 10/01/2021	Planned close out: 09/30/2024
Project Finish Date	Planned Early Finish:	Estimated Early Finish:
Project %-complete	35%	

## 2. Near-Term Milestones

Include milestones with the scheduled dates or actual/forecast dates that are in current and the next reporting period, and milestones (with past scheduled dates) that are delayed to future reporting period. (**Completed deliverables have bold font dates.**)

WB S	Subsystem	Milestone	Scheduled Date	Actual date (A) /Forecast Date (F)
1.1	1.1.2.2	<b>Complete first draft of data needs report (based on 1.1.1 -M6 milestone)</b>	02/28/2023	<b>02/28/2023(A)</b>
	1.1.2.4	<b>Database of Peers (RV) created</b>	10/31/2022	<b>10/31/2022(A)</b>
	1.1.3	Monitors Requirements documented, hardware and software needs assessed	9.30/2023	9/30/2023(F)
	1.1.3.3	Telescope data monitoring needs compiled	05/31/2023	05/31/2023(F)
	1.3.3.4	Two-way traffic data monitoring needs compiled	05/31/2023	05/31/2023(F)
	1.3.3.5	BGP data monitoring needs compiled	09/30/2023	09/30/2023(F)
	1.1.3.6	Active measurements needs compiled	03/31/2023	05/31/2023(F)
	1.1.3.7	DNS monitoring needs compiled	04/30/2023	04/30/2023(F)
	1.1.4.1	Monitor specification report draft posted for internal feedback	03/31/2023	05/31/2023(F)
	1.1.4.2	Draft (2) for community feedback	05/31/2023	07/31/2023(F)
	1.1.4.3	Post final report for public comments	09/30/2023	09/30/2023(F)
	1.1.5.1	<b>Telescope data monitor software prototyped</b>	03/31/2023	<b>03/31/2023(A)</b>
	1.1.5.2	<b>Two-way traffic data monitor software prototyped</b>	01/31/2023	<b>01/31/2023(A)</b>
	1.1.5.3	<b>BGP data monitor software prototyped</b>	2/28/2023	<b>2/28/2023(A)</b>
	1.1.5.4	<b>Active data monitor software prototyped</b>	2/28/2023	<b>2/28/2023(A)</b>
	1.1.6.2	Telescope data monitoring deployment	07/01/2023	07/01/2023(F)
	1.1.6.3	Two-way traffic monitor deployment	1/31/2023	06/30/2023(F)
	1.1.6.5	<b>Active probing measurements data monitoring deployment</b>	2/28/2023	<b>2/28/2023(A)</b>
	1.1.6.6	DNS data monitoring deployment	05/31/2023	05/31/2023(F)
	1.1.6.7	<b>10 nodes of at least one measurement deployed</b>	3/31/2023	<b>3/31/2023(A)</b>
	1.1.6.8	10 nodes of multiple measurements deployed	09/30/2023	09/30/2023(F)
	1.1.7.1	<b>Software to support active probing measurements deployed</b>	9/30/2022	<b>3/31/2022(A)</b>
	1.1.9	Integrated report "Internet infrastructure security vulnerabilities" and "Data Needs" published	2/28/2023	<b>5/01/2023 (A)</b>
1.2	1.2.1	Data storage hardware requirement documented - draft posted for stakeholders' review	3/31/2023	6/30/2023(F)
	1.2.2	Data storage systems specifications published	7/01/2023	7/01/2023(F)
	1.2.2.2	Data storage systems specifications documented- <b>draft posted for stakeholders' review</b>	2/28/2023	6/30/2023(F)
	1.2.2.3	Community feedback incorporated into document	07/01/2023	07/01/2023(F)
	1.2.3.1.2	<b>Create metadata templates for other ongoing data sets</b>	3/31/2023	<b>3/31/2023(A)</b>
	1.2.3.1.3	<b>Create metadata templates for the datasets</b>	10/31/2022	<b>10/31/2022(A)</b>

		<b>completed in the last 5 years</b>		
	<b>1.2.3.2</b>	<b>Research the state-of-the-art metadata approaches</b>	10/31/2022	<b>10/31/2022(A)</b>
	<b>1.2.3.4</b>	<b>YR1 Data and metadata standards specifications published</b>	10/31/2022	<b>03-31-2023(A)</b>
	1.2.3.4-2	YR2 Data and metadata standards specifications published	9/30/2023	09-30-2023(F)
	1.2.4.2	Report on pcap real-time analysis tools	7/01/2023	7/01/2023(F)
	<b>1.2.4.3</b>	<b>Specification of tools for data curation and documentation, YR1 report created</b>	3/31/2023	<b>3/31/2023(A)</b>
	<b>1.2.5.1</b>	<b>Unified web interface to download heterogeneous datasets designed</b>	3/31/2023	<b>3/31/2023 (A)</b>
	1.2.5.2.1	Incorporate datasets used by IYP	07/01/2023	07/01/2023(F)
	<b>1.2.5.3.</b>	<b>Improve existing API</b>	2/28/2023	<b>2/28/2023(A)</b>
	<b>1.2.5.4</b>	<b>Document data and metadata APIs</b>	3/31/2023	<b>3/31/2023(A)</b>
	1.2.5.4.2	Update of data and metadata API	09/30/2023	09/30/2023(F)
	1.2.8	Approaches to data dissemination designed, documented	09/30/2023	09/30/2023(F)
	1.2.8.1	Report on the latest big data storage/management technologies	3/31/2023	<b>06/30/2023(F)</b>
1.3				
	1.3.1	Data discovery tools prototyped	07/01/2023	07/01/2023
	<b>1.3.1.3</b>	<b>Prototype GMI3S Data Catalog</b>	10/31/2022	<b>03/31/2022(A)</b>
	<b>1.3.1.2</b>	<b>Documentations of the existing tools and datasets improved</b>	10/31/2022	<b>10/31/2022(A)</b>
	1.3.1.4	Report on other non-CAIDA datasets and tools integration	05/31/2023	05/31/2023(F)
	1.3.1.5	Report on integration of automated meta-data/data citation creation into catalog	07/01/2023	07/01/2023(F)
	<b>1.3.1.6</b>	<b>Metadata databases to increase data accessibility created</b>	1/31/2023	<b>2/28/2023(A)</b>
	1.3.2.1	Report on the gaps between privacy-preservation techniques and network and security research needs	12/31/2022	<b>09/30/2023(F)</b>
	1.3.2.3	Gaps that privacy techniques can support identified, report created and shared	1/31/2023	<b>09/30/2023(F)</b>
	1.3.2.4	Create taxonomy of data	12/31/2022	<b>09/30/2023(F)</b>
	<b>1.3.2.6</b>	<b>Design and prototype authentication/authorization solution</b>	12/31/2022	<b>01/31/2023(A)</b>
	<b>1.3.3.1</b>	<b>A list of existing data sharing policies and best practices compiled and shared with community</b>	3/31/2023	<b>3/31/2023(A)</b>
	<b>1.3.3.5</b>	<b>Lessons learned (from previous data-sharing efforts) documented</b>	3/31/2023	<b>03/31/2023(A)</b>
	1.3.3.6	Report on other countries' approaches to policy tools for disclosure control	3/31/2023	<b>09/30/2023(F)</b>
	<b>1.3.4.4</b>	<b>Case study comparing modes of sharing the darknet dataset with three entities</b>	3/31/2023	<b>3/31/2023(A)</b>

		<b>documented and shared</b>		
1.4	1.4.1	<b>GMI quarterly workshop conducted</b>	03/31/2023	<b>03/31/2023(A)</b>
	<b>1.4.2.2</b>	<b>Virtual collaboration environment evaluated and improved</b>	09/30/2022	<b>09/30/2022(A)</b>
	<b>1.4.3.1</b>	<b>Modules to scale STEM workforce developed</b>	09/30/2023	<b>03/31/2023(A)</b>
	1.4.3.2	Online course on NIDS developed	09/30/2023	09/30/2023(F)
	1.4.3.3	Video tutorials on nodes deployment and management created	3/31/2023	09/30/2023(F)
	<b>1.4.3.4</b>	<b>Quarterly calls conducted, minutes shared</b>	03/31/2023	<b>03/31/2023(A)</b>
	<b>1.4.3.5</b>	<b>Project Presentations</b>	03/31/2023	<b>03/31/2023(A)</b>

### 3. Technical progress highlights

This section summarizes highlights of progress of the current period, by near-term tasks, including the following select achievements:

- We incorporated another few rounds of feedback into our “[Internet vulnerabilities and data needs report](#)”, and lessons learned from previous data sharing efforts, and published it on our website. As part of this effort, we engaged with additional stakeholders from industry, academia, and government to gain insights into measurement needs and data acquisition infrastructure design. Perhaps most importantly, we initiated a series of meetings (which continues without us) with NIST and NTIA to pursue US government participation in best practices for routing security (i.e., RPKI, anti-spoofing) as an alternative/complement to the FCC’s impending threat of regulation of U.S.-based networks participating in the BGP routing system. **(Milestone 1.1.9, 1.3.3.5)**
- Clark, Testart (MIT), and Claffy (UCSD) submitted a public comment to the [FCC’s Notice of inquiry](#) regarding Internet routing security which the FCC [published in April 2022](#) .
- We worked on a follow up document offering additional perspectives on these issues. We summarized and elucidated discussion topics and compiled related references in the [document](#) published by the FCC on Jan 24, 2023.
- We established a new MOU with DomainTools and two service agreements with academic and government stakeholders to support telescope traffic data sharing. **(Milestone 1.3.3.1)**
- We established three monthly call series on Darknet Traffic, DDOS, and active measurement to discuss requirements for these types of measurements. **(Milestones 1.4.1, 1.4.4)**
- We are leading weekly calls on a DNS security topic as part of ICANN’s industry-led Security Advisory Committee. **(Milestone 1.4.3.4)**
- We prototyped and documented software components for Telescope (darknet) traffic, two-way traffic, BGP and active-measurement monitors. (Milestones 1.1.5.[1-4])
- We prototyped an inventory system for CAIDA’s tools and data. **(Milestone 1.1.3.1 & 1.1.3.2).**
- We discussed the BGP data infrastructure vision, goals, stats and challenges during the project retreat in January 2023. **(Milestone 1.1.3.5)**
- We executed Ark-related infrastructure transition, due to the retirement of the project lead. This was painful but informative, as we identified many pain points in the current design, redesigned software components, and deployed them on our own and other infrastructures, including industry (ExpressVPN) and academic (RouteViews) **(Milestone 1.1.3.6, 1.1.5.4, 1.1.6.[5-8])**
- We documented and made updates to [CAIDA data APIs](#), including a new [geolocation API based on DNS hostnames](#). **(Milestones 1.2.5)**
- We used the DNS Top Level Zone Database (<https://dzdb.caida.org>) to experiment with data curation to support security use of the data. We modified the API to let users type in a prefix and get all nameserver IPs/hostnames and domains served by that prefix. One implication of this functionality is that we could take blacklist data and study 'how many domains are served by these blacklisted IPs'.
- RV has been working with Google to incorporate their data into Google’s BigQuery. CAIDA participated in a series of meetings. Those included a BQ training session where we learned the BGP data schema and the way to query the data.
- We completed the first draft of our [Metadata Annotated Schema](#) to facilitate curation of GMI data sets **(Milestone 1.2.3, 1.2.4)**
- Study of compliance with DNS security best practices to ascertain DNS measurement needs: “[Observable KINDNS: Validating DNS Hygiene](#)”. **(Milestone 1.1.5.5)**
- We published a working version of D. Clark’s note “[The EU NIS-2 proposal and the DNS](#)” that summarizes the potential impact on the DNS **(Milestone 1.1.3.7)**
- We deployed more than ten nodes of active measurements data monitors on various VPs , and deployed software to support these measurements **(Milestone 1.1.6.7, 1.1.7.1)**

- We submitted a paper “[A path forward: Improving Internet routing security by enabling trust zones](#)” that demonstrates how systematic data collection can achieve progress against the most persistent and harmful vulnerabilities in the routing system.
- We decided to shift from OpenStack to Ceph/MinIO for storage, and engaged with the Open Science Data Federation project to explore its use for GMI. **(Milestone 1.2.2)**
- Completed prototype of SSO authentication and authorization framework. **(Milestone 1.3.2.6)**
- Completed first design and prototype of [BGP data science gateway](#). **(Milestone 1.2.5.1)**
- CAIDA’s [Periscope looking glass API](#) provides a uniform interface to hundreds of Looking Glass and other network vantage points that perform traceroute and BGP queries. We enhanced this API with new features that include “scamper-routeviews” (periscope now implements the "traceroute" command on scamper processes running on Routeviews collectors), new monitor “type” and “server” fields, and new Keycloak-based authentication/authorization framework.
- The [Resource Catalog](#) provides a unified interface to metadata and relationships between datasets, papers, presentations, media, software, and recipes (code and instructions on how to solve various Internet security-related problems using datasets, tools and other objects indexed in catalog).
- We conducted a case study comparing and documenting the darknet dataset use by three entities **(Milestone 1.3.4.4)**
- We are now experimenting with sharing Telescope data in several different ways with researchers.
- We designed and taught [Internet Data Science for Cybersecurity](#) at UCSD **(Milestone 1.4.3)**
- All materials related to the “[Internet Data Science for Cybersecurity](#)” course are now available online and in github repositories and can be used for self-paced or in-class studies.
- We discussed the RV project vision, goals, stats and challenges during the project retreat in January 2023
- We held our first Strategic Advisory Council meeting, hereafter quarterly. **(Milestone 1.5.3.2)**
- PI Claffy co-hosted Telecommunications Policy Research Conference (TPRC) graduate students’ workshops in DC. The goals of the Graduate Student Workshop are to orient graduate students to the TPRC conference and community ahead of the conference, to provide feedback and actionable advice that will improve the quality and effective communication of student research, and to foster networking between early career graduate students, more senior graduate students, and mentors from academia and industry.
- CAIDA researchers co-authored five papers and 3 posters that were presented at ACM’s IMC 2022 conference in Nice, France on October 25 – 27, 2022. For more details see our blog at [https://blog.caida.org/best\\_available\\_data/2022/10/18/caida-contributions-to-acms-internet-measurement-conference-imc-2022/](https://blog.caida.org/best_available_data/2022/10/18/caida-contributions-to-acms-internet-measurement-conference-imc-2022/)

## 4. List of Acronyms

AIMS	Active Internet Measurements
AMPRNet	AMateur Packet Radio Network (Network 44)
API	Application Programming Interface
Ark	CAIDA Archipelago Measurement Infrastructure
AS	Autonomous System
AUA	Acceptable Use Agreement
AUP	Acceptable Use Policy
AY	At Year
BCP	Baseline Change Proposal
BGP	Border Gateway Protocol
BMP	BGP Monitoring Protocol
CAIDA	The Center for Applied Internet Data Analysis
CDN	Content Delivery Network
CDR	Conceptual Design Report
CENIC	California Educational & Research Network Information Center
CFR	Code of Federal Regulations
CI	Cyber Infrastructure
CISE	Computer and Information Science Engineering
CNS	Computer and Networked Systems
CMP	Configuration Management Plan
CPI	Cost Performance Index
CSAIL	Computer Science & Artificial Intelligence Laboratory
CY	Calendar Year
DDOS	Distributed Denial of Service
DHS	Department of Homeland Security
DNS	Domain Name System
DPM	Deputy PM
DPU	Data Processing Unit
DREN	Defence Research and Engineering Network
DZDB	DNS Zone Database
EAC	Estimate at Completion
eMMC	embedded Multi Media Card
ES&H	Environment, Safety and Health
EVMS	Earned Value Management System
FCC	Federal Communication Commissions

FFRDC	Federally Funded Research and Development Center
FTE	Full Time Equivalent Employee
FY	Fiscal Year
GE	Gigabit Ethernet
GMI3S	Global Measurement Infrastructure to Improve Internet Security
ICANN	Internet Corporation for Assigned Names and Numbers
IHR	Internet JHealth Report
IJ	Internet Initiative Japan
IMC	Internet Measurement Conference
IOT	Internet of Things
ISP	Internet Service Provider
IYP	Internet Yellow Pages (II)
KINDNS	Knowledge-Sharing and Instantiating Norms for DNS and Naming Security
L2	Level 2
LFO	Large Facilities Office
LoC	Letter of Collaboration
MANRS	Mutually Agreed Norms for Routing Security
MIT	Massachusetts Institute of Technology
MREFC	Major Research Equipment and Facilities Construction
NIC	Network Interface Card
NIDS	Network Infrastructure Data Science
NIS	Network and Information Security (EU)
NIST	National Institute of Standards and Technology
NITRD	Network and Information Technology Research and Development
NLP	Natural Language Processing
NOG	Internet Network Operators' Group
NSF	National Science Foundation
NSRC	Network Startup Resource Center
NTIA	National Telecommunications and Information Administration
OAC	Office of Advance Cyberinfrastructure
OMB	Office of Management and Budget
PI	Principal Investigator
PII	Personally Identifiable Information
PEP	Project Execution Plan
PM	Project Manager
PMCS	Project Management Control System
PO	Program Officer 5
PRP/NRP	Pacific Research Platform/National Research Platform

QA	Quality Assurance
QC	Quality Control
R&D	Research and Development
RIPE-NCC	Regional Internet Registry for Europe Network Coordination Centre
RIS	Routing Information System
RLS	Resource Loaded Schedule
ROA	Routing Origin Authorization



RPKI	Resource Public Key Infrastructure
RV	Route Views
SD card	Secure Digital card
SDSC	San Diego Supercomputer Center
SOW	Statement of Work
SPI	Schedule Performance Index
SSAC	Security and Stability Advisory Committee
TLD	Top-level domain
TPC	Total Project Cost
UO	University of Oregon
U.S.	United States
USC ISI	USC Information Science Institute
VP	Vantage Point
WBS	Work Breakdown Structure