Impact
Cyber Trust

Information Marketplace for Policy and Analysis of Cyber-risk & Trust

# Driving Data in the Cybersecurity Economy

The Economist

JULY 6TH–12TH 2017

Crunch time in France

Ten years on: banking after the crisis

South Korea's unfinished revolution

Biology, but without the cells

The world's most valuable resource

Data and the new rules of competition

Homeland Security
Science and Technology

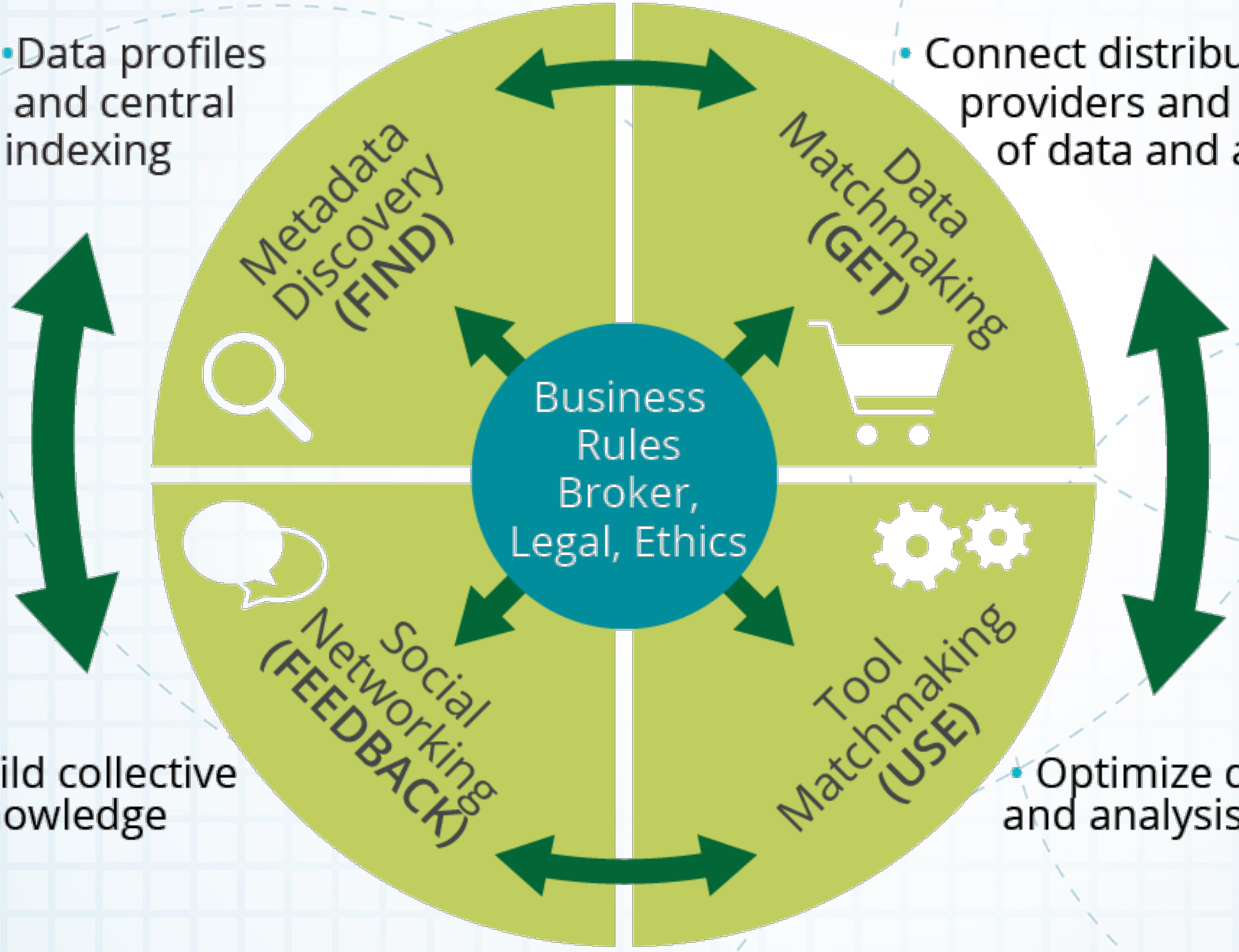
**Erin Kenneally**
**U.S. Dept of Homeland Security**
**Cyber Security Division**

# IMPACT Motivation: The ‘Open Secret’ of Effective R&D

- **Data are critical to R&D capabilities**
  - Exactly 0% of R&D (quality) possible sans data
  - Cybersecurity needs real-world data to develop, test, evaluate knowledge & tech solutions to counter cyber threats
  - “Big Data” may grow on trees but still has to be picked, sorted, trucked

- **Decision analytics are critical to HSE capabilities**
  - Cybersecurity needs integrated, holistic understanding of risk environment
  - Gap between Data <-->Decisions: multi-dimensional, complex association and fusion, high-context presentation elements

- **Data sharing + Analytics |= Easy**
  - High value data = High legal risk + $$
  - Data rich vs. data poor
  - Expensive to abstract away low level knowledge- and labor- intensive tasks
  - Technologists optimize for Efficiency, Lawyers optimize for Certainty


Metadata Discovery (FIND)
Data Matchmaking (GET)
Business Rules Broker, Legal, Ethics
Social Networking (FEEDBACK)
Tool Matchmaking (USE)

- Data profiles and central indexing
- Connect distributed providers and seekers of data and analytics
- Build collective knowledge
- Optimize data use and analysis

2018 Kenneally

# IMPACT ROI

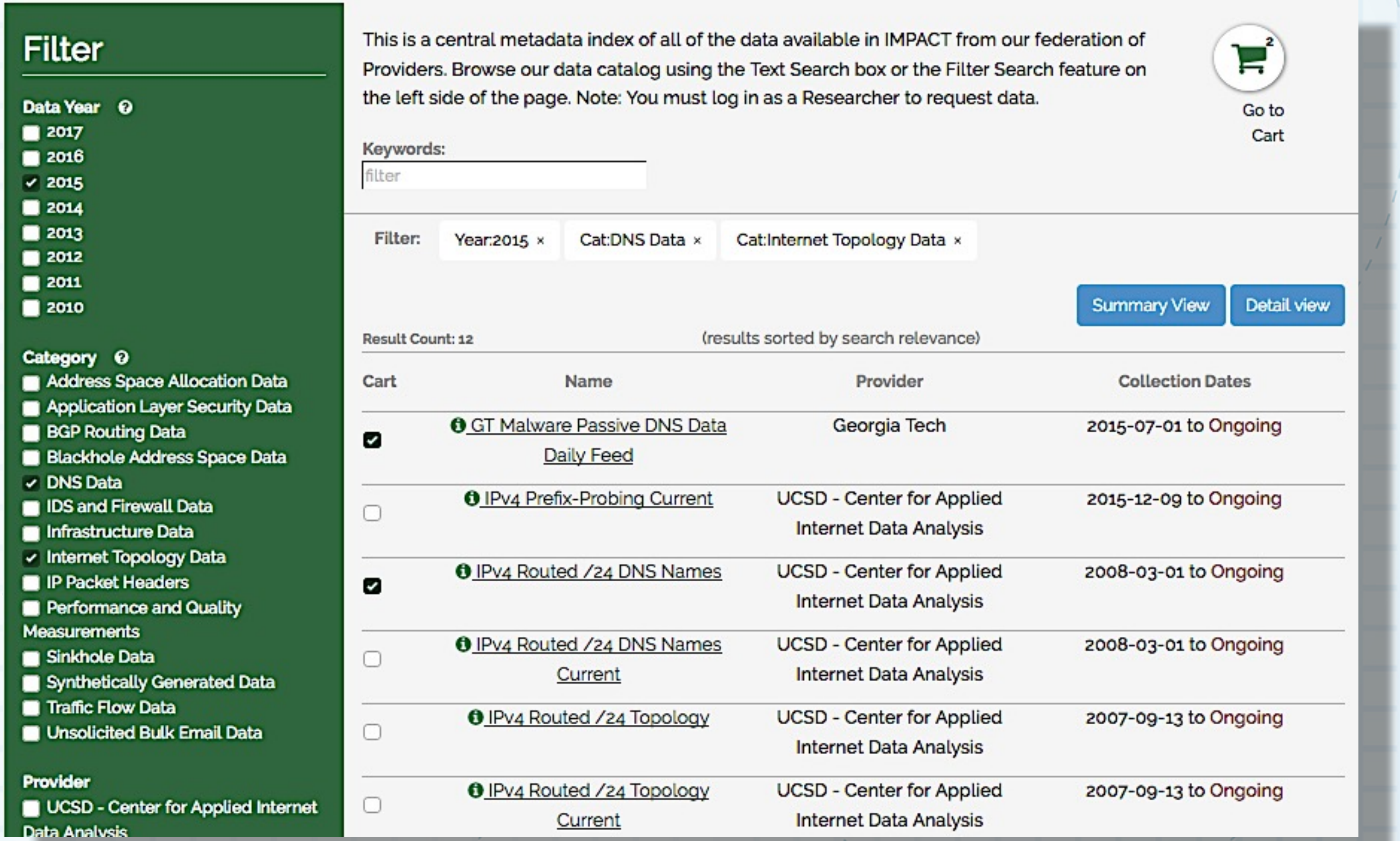
- **Parity-** lower barrier to entry for data impoverished viz federation of data Supply & Demand (academic, industry, govt)

- **Scale-** beyond interpersonal relationships, ad hoc acquisitions
- **Sustainable-** Uniform, repeatable process

- **Utility-** responsible innovation over risk-aversion

- **Trust**
  - Vetted data, researchers, providers
  - Balance efficiency and certainty
  - Legal and ethical accountability

# Shop til You Drop
# IMPACT Portal <www.ImpactCyberTrust.org>



## Filter

**Data Year** ?

- □ 2017
- □ 2016
- ☑ 2015
- □ 2014
- □ 2013
- □ 2012
- □ 2011
- □ 2010

**Category** ?

- □ Address Space Allocation Data
- □ Application Layer Security Data
- □ BGP Routing Data
- □ Blackhole Address Space Data
- ☑ DNS Data
- □ IDS and Firewall Data
- □ Infrastructure Data
- ☑ Internet Topology Data
- □ IP Packet Headers
- □ Performance and Quality Measurements
- □ Sinkhole Data
- □ Synthetically Generated Data
- □ Traffic Flow Data
- □ Unsolicited Bulk Email Data

**Provider**

- □ UCSD - Center for Applied Internet Data Analysis

This is a central metadata index of all of the data available in IMPACT from our federation of Providers. Browse our data catalog using the Text Search box or the Filter Search feature on the left side of the page. Note: You must log in as a Researcher to request data.

2 Go to Cart

Keywords: filter

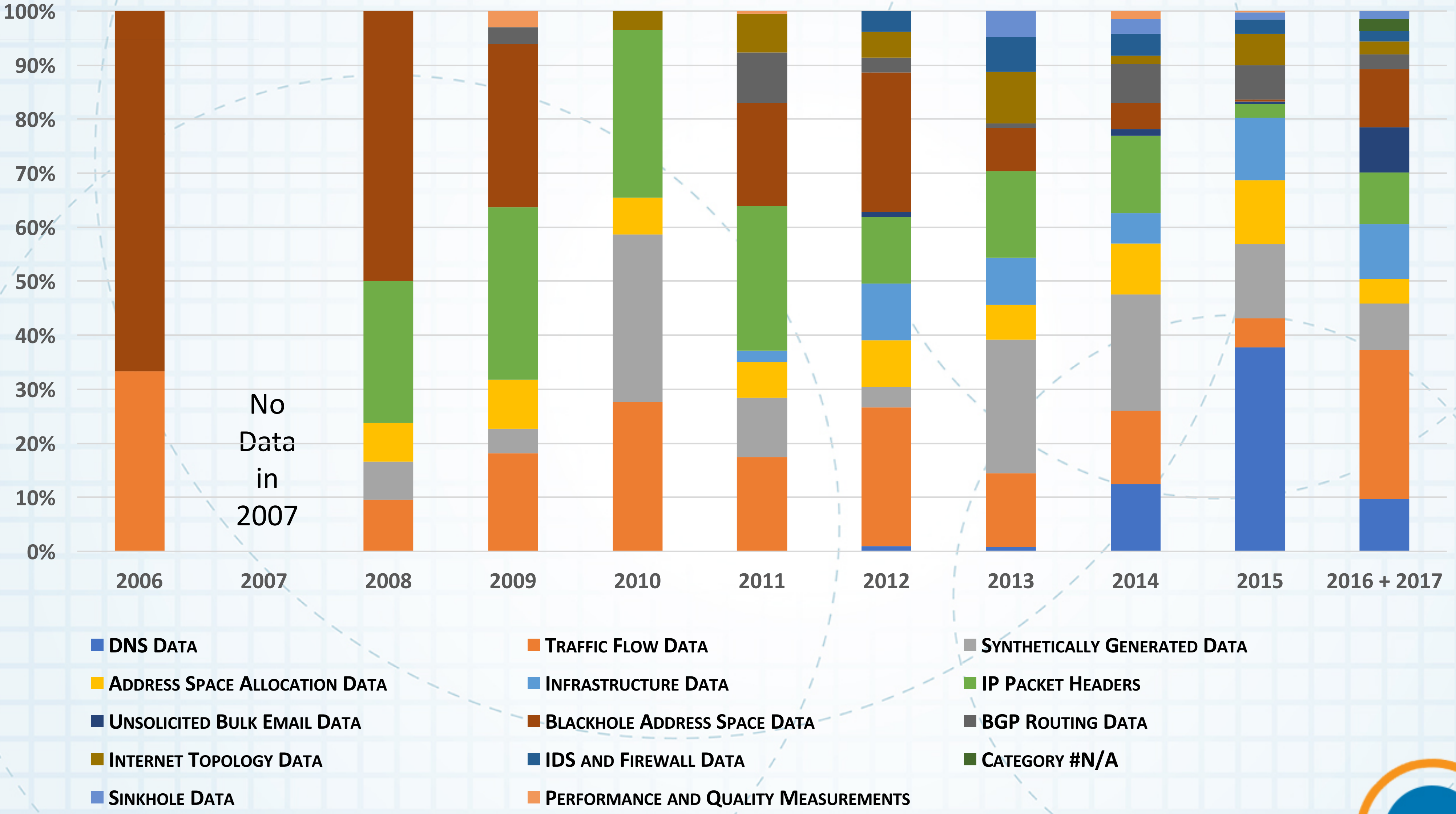
Filter: Year:2015 × Cat:DNS Data × Cat:Internet Topology Data ×

Summary View Detail view

Result Count: 12 (results sorted by search relevance)

| Cart | Name | Provider | Collection Dates |
|---|---|---|---|
| ☑ | GT Malware Passive DNS Data Daily Feed | Georgia Tech | 2015-07-01 to Ongoing |
| □ | IPv4 Prefix-Probing Current | UCSD - Center for Applied Internet Data Analysis | 2015-12-09 to Ongoing |
| ☑ | IPv4 Routed /24 DNS Names | UCSD - Center for Applied Internet Data Analysis | 2008-03-01 to Ongoing |
| □ | IPv4 Routed /24 DNS Names Current | UCSD - Center for Applied Internet Data Analysis | 2008-03-01 to Ongoing |
| □ | IPv4 Routed /24 Topology | UCSD - Center for Applied Internet Data Analysis | 2007-09-13 to Ongoing |
| □ | IPv4 Routed /24 Topology Current | UCSD - Center for Applied Internet Data Analysis | 2007-09-13 to Ongoing |

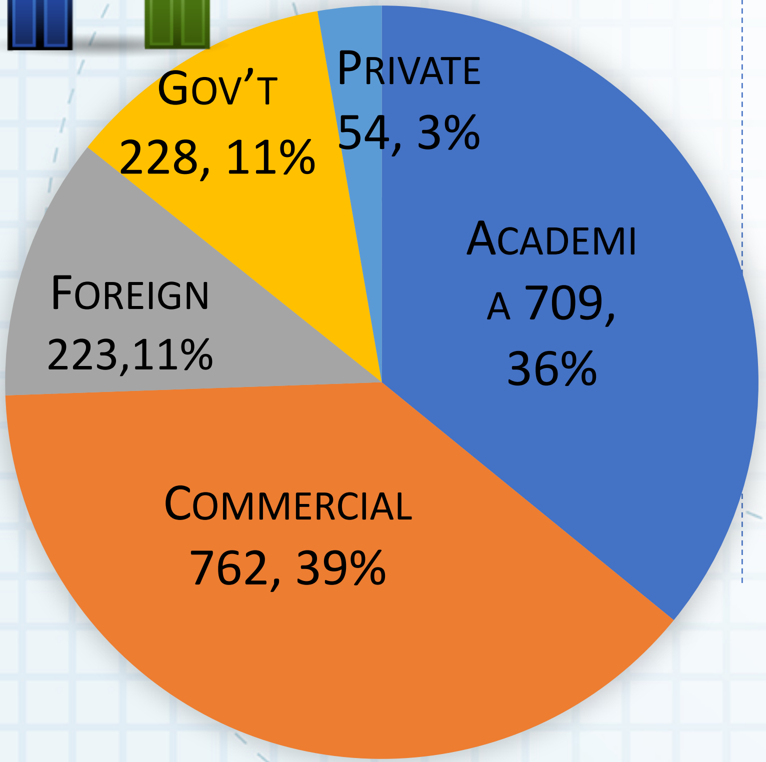
# Data Trends



Source: DHS IMPACT program; SRI analysis, Apr '17

# Global, Multi-Sector "Impact" (as of Jul 2017)

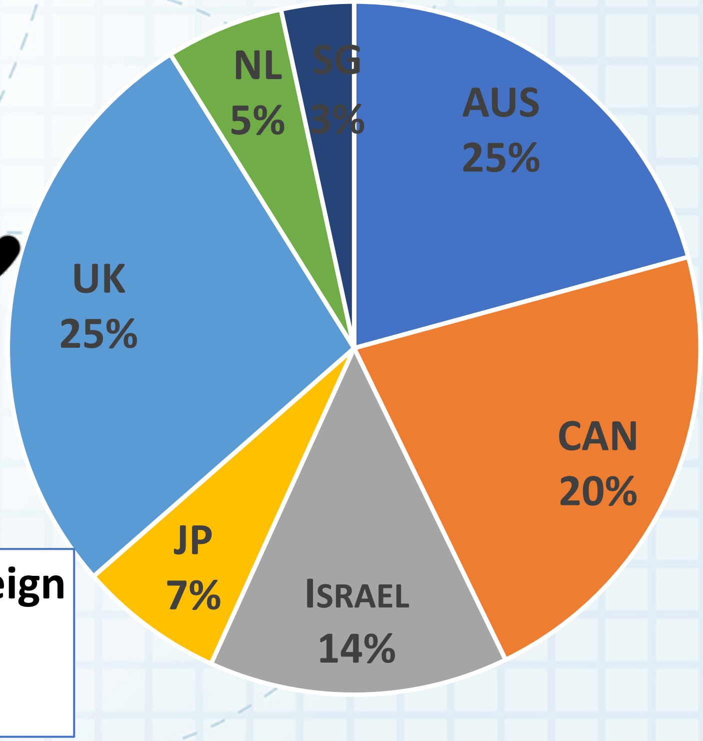
Total Users
(1,987)



Dataset Provisioned
(>3,500)

Research papers, journals,
tech reports (>300 "known")

Approved Foreign
Users
(236 Total)



Source: DHS IMPACT program; SRI analysis, Apr '17

Impact Cyber Trust

# Success Elements

**Findable**

Centralized Mediation

**Tools to USE the data**

**Diverse Real-world** Problem-driven Data

**Responsible** Legal & Ethical framework integrated

New, **high-value** datasets

**FREE**

Engage **International** data and researchers

**Distributed Provisioning**

2018 Kenneally

# Evolved IMPACT R&D Approach

## Market need:

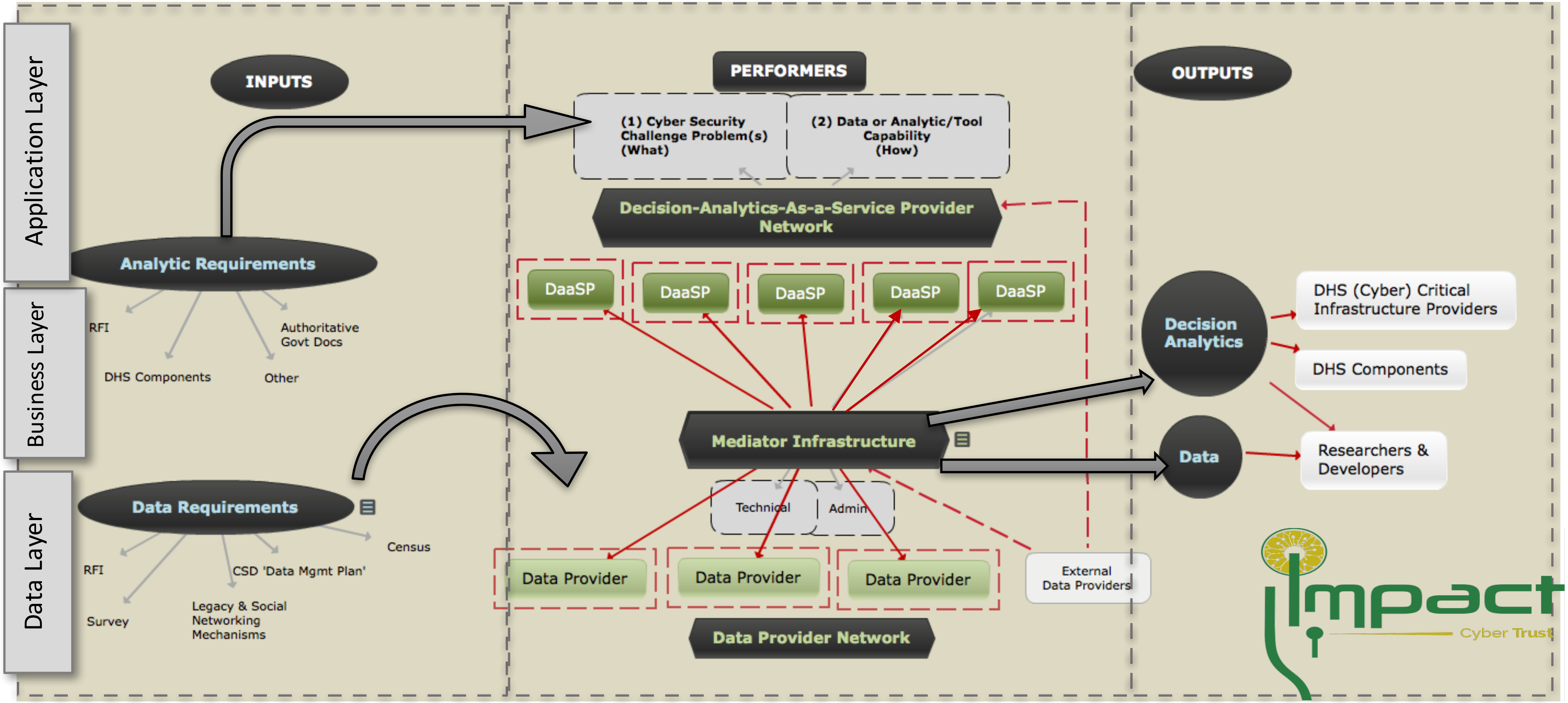
- Existing capabilities do not provide cyber risk decision analytic support needed by HSE
  - **Security, Integrity, Stability, Resilience of networks**
  - **Sensitive data sharing and controlled data disclosure**
  - **Interdependencies, cascading, and aggregate effects of cyber-vulnerabilities and attacks across platforms and enterprises**
- Changing risk environment demands dynamic cyber security R&D
- < time & effort to find, curate, normalize, understand high volume, velocity, variety, value
  > time extracting insight and meaningful decisions from data

## Product:

- 1st-gen R&D-enabling infrastructure democratized ***data raw materials*** **(Data Providers)**
- New BAA fosters evolved R&D infrastructure adds ***derivative data products and tools*** for HSE: **Decision Analytics-as-a-Service Providers (DASP)**

2018 Kenneally

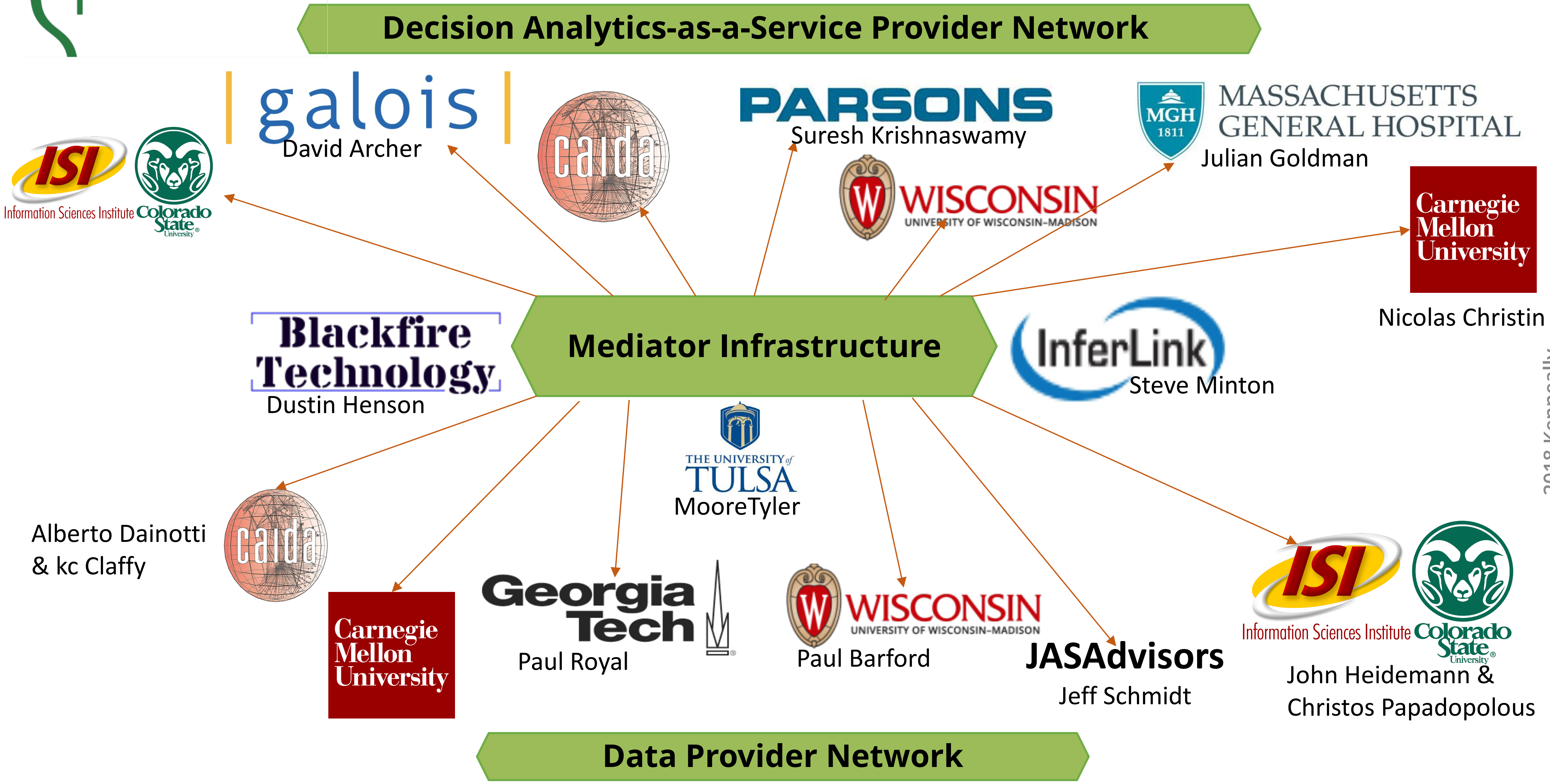
# NGI Recap



2018 Kenneally

Impact Cyber Trust

# Class of 2018



Decision Analytics-as-a-Service Provider Network

Mediator Infrastructure

Data Provider Network

galois — David Archer

caida

PARSONS — Suresh Krishnaswamy

WISCONSIN UNIVERSITY OF WISCONSIN–MADISON

MGH 1811 MASSACHUSETTS GENERAL HOSPITAL — Julian Goldman

ISI Information Sciences Institute, Colorado State University

Carnegie Mellon University — Nicolas Christin

Blackfire Technology — Dustin Henson

InferLink — Steve Minton

THE UNIVERSITY of TULSA — MooreTyler

caida — Alberto Dainotti & kc Claffy

Carnegie Mellon University

Georgia Tech — Paul Royal

WISCONSIN UNIVERSITY OF WISCONSIN–MADISON — Paul Barford

JASAdvisors — Jeff Schmidt

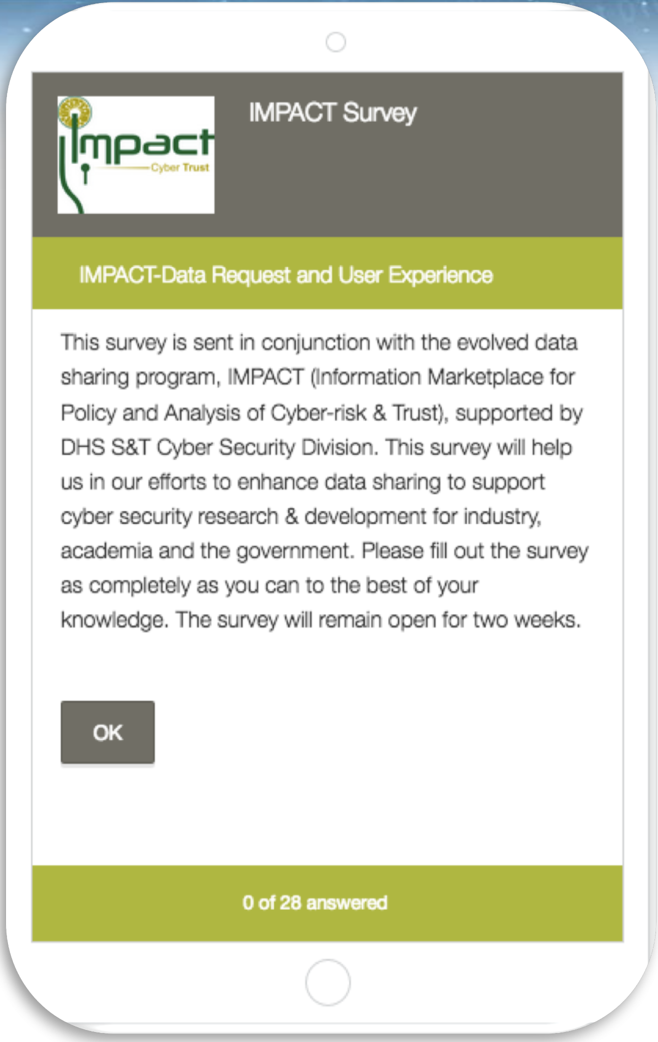
ISI Information Sciences Institute, Colorado State University — John Heidemann & Christos Papadopolous

2018 Kenneally

# Socialization

impact Cyber Trust

Annual Digest
November 2017



impact Cyber Trust

IMPACT Survey

IMPACT-Data Request and User Experience

This survey is sent in conjunction with the evolved data sharing program, IMPACT (Information Marketplace for Policy and Analysis of Cyber-risk & Trust), supported by DHS S&T Cyber Security Division. This survey will help us in our efforts to enhance data sharing to support cyber security research & development for industry, academia and the government. Please fill out the survey as completely as you can to the best of your knowledge. The survey will remain open for two weeks.

0 of 28 answered

impact

Information Marketplace for Policy and Analysis of Cyber-risk & Trust

PRESS KIT | 2017

https://www.ImpactCyberTrust.org/#knowledgebase

# Why Engage IMPACT

- **How do companies address risks associated with data sharing for academic research?***

  - Engage in a rigorous internal review of proposed academic research projects.
  - Close to half of the companies retain custody and control over the research data at all times.
  - Companies employ rigorous data use agreements to limit access to and use of shared data.

**How IMPACT addresses risks**

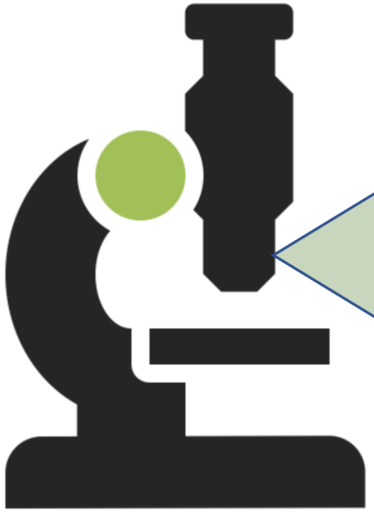
- Vet Researchers, Providers, Data
- Provider can host and provision own data
- Provider can engage Disclosure Control-as-a-Service for very sensitive data that allows analysis without Researcher seeing data
- Provider leverages standardized Researcher data use agreements with customized additional restrictions by Provider

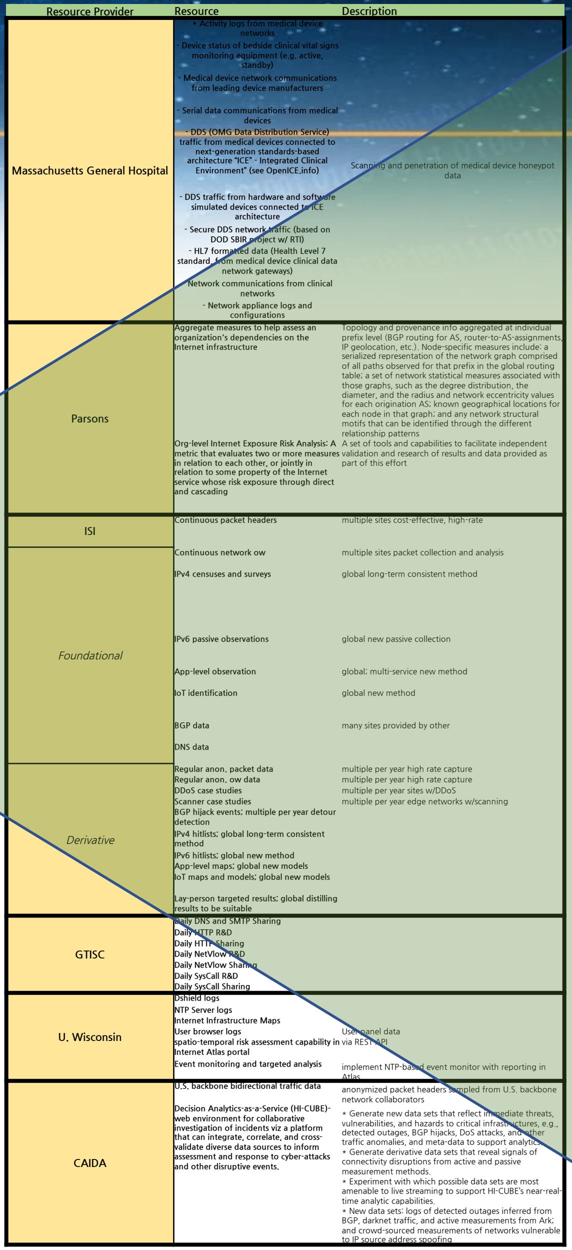
* "UNDERSTANDING CORPORATE DATA SHARING DECISIONS:PRACTICES, CHALLENGES, AND OPPORTUNITIES FOR SHARING CORPORATE DATA WITH RESEARCHERS" Future of Privacy Forum (2017)

# Popularity

| Name | Data Provider |
|---|---|
| GT Malware Passive DNS Data Daily Feed | Georgia Tech |
| Historical GT Malware Passive DNS Data 2011-2013 | Georgia Tech |
| US Long-haul Infrastructure Topology | University of Wisconsin |
| DARPA Scalable Network Monitoring (SNM) Program Traffic | DARPA |
| Skaion 2006 IARPA Dataset | SKAION |
| GT Malware Unsolicited Email Daily Feed | Georgia Tech |
| DSHIELD Logs | University of Wisconsin |
| syn-flood-attack | Merit Network, Inc. |
| Netflow-1 | Merit Network, Inc. |
| DoS_traces-20020629 | University of Southern California-Information Sciences Institute |
| NCCDC 2013 | Center for Infrastructure Assurance and Security (UTSA/CIAS) |
| NCCDC 2014 | Center for Infrastructure Assurance and Security (UTSA/CIAS) |
| DoS_80_timeseries-20020629 | University of Southern California-Information Sciences Institute |
| CAIDA DDoS 2007 Attack Dataset | UCSD - Center for Applied Internet Data Analysis |
| Netflow-2 | Merit Network, Inc. |
| Netflow-3 | Merit Network, Inc. |
| NCCDC 2011 | Center for Infrastructure Assurance and Security (UTSA/CIAS) |
| NTP DDoS 2014 | Merit Network, Inc. |
| NCCDC 2015 | Center for Infrastructure Assurance and Security (UTSA/CIAS) |
| UCSD Real-time Network Telescope Data | UCSD - Center for Applied Internet Data Analysis |

# Booths and Wares in the Marketplace:





| Resource Provider | Resource | Description |
|---|---|---|
| Massachusetts General Hospital | * Activity logs from medical device networks<br>- Device status of bedside clinical vital signs monitoring equipment (e.g. active, standby)<br>- Medical device network communications from leading device manufacturers<br>- Serial data communications from medical devices<br>- DDS (OMG Data Distribution Service) traffic from medical devices connected to next-generation standards-based architecture "ICE" - Integrated Clinical Environment" (see OpenICE.info)<br>- DDS traffic from hardware and software simulated devices connected to ICE architecture<br>- Secure DDS network traffic (based on DOD SBIR project w/ RTI)<br>- HL7 formatted data (Health Level 7 standard, from medical device clinical data network gateways)<br>Network communications from clinical networks<br>- Network appliance logs and configurations | Scanning and penetration of medical device honeypot data |
| Parsons | Aggregate measures to help assess an organization's dependencies on the Internet infrastructure | Topology and provenance info aggregated at individual prefix level (BGP routing for AS, router-to-AS-assignments, IP geolocation, etc.). Node-specific measures include: a serialized representation of the network graph comprised of all paths observed for that prefix in the global routing table; a set of network statistical measures associated with those graphs, such as the degree distribution, the diameter, and the radius and network eccentricity values for each origination AS; known geographical locations for each node in that graph; and any network structural motifs that can be identified through the different relationship patterns |
| | Org-level Internet Exposure Risk Analysis: A metric that evaluates two or more measures in relation to each other, or jointly in relation to some property of the Internet service whose risk exposure through direct and cascading | A set of tools and capabilities to facilitate independent validation and research of results and data provided as part of this effort |
| ISI | Continuous packet headers | multiple sites cost-effective, high-rate |
| *Foundational* | Continuous network ow | multiple sites packet collection and analysis |
| | IPv4 censuses and surveys | global long-term consistent method |
| | IPv6 passive observations | global new passive collection |
| | App-level observation | global; multi-service new method |
| | IoT identification | global new method |
| | BGP data | many sites provided by other |
| | DNS data | |
| *Derivative* | Regular anon. packet data | multiple per year high rate capture |
| | Regular anon. ow data | multiple per year high rate capture |
| | DDoS case studies | multiple per year sites w/DDoS |
| | Scanner case studies | multiple per year edge networks w/scanning |
| | BGP hijack events; multiple per year detour detection | |
| | IPv4 hitlists; global long-term consistent method | |
| | IPv6 hitlists; global new method | |
| | App-level maps; global new models | |
| | IoT maps and models; global new models | |
| | Lay-person targeted results; global distilling results to be suitable | |
| GTISC | Daily DNS and SMTP Sharing<br>Daily HTTP R&D<br>Daily HTTP Sharing<br>Daily NetVlow R&D<br>Daily NetVlow Sharing<br>Daily SysCall R&D<br>Daily SysCall Sharing | |
| U. Wisconsin | Dshield logs | |
| | NTP Server logs | |
| | Internet Infrastructure Maps | |
| | User browser logs | User panel data |
| | spatio-temporal risk assessment capability in Internet Atlas portal | via REST API |
| | Event monitoring and targeted analysis | implement NTP-based event monitor with reporting in Atlas |
| CAIDA | U.S. backbone bidirectional traffic data | anonymized packet headers sampled from U.S. backbone network collaborators |
| | Decision Analytics-as-a-Service (HI-CUBE)-web environment for collaborative investigation of incidents viz a platform that can integrate, correlate, and cross-validate diverse data sources to inform assessment and response to cyber-attacks and other disruptive events. | * Generate new data sets that reflect immediate threats, vulnerabilities, and hazards to critical infrastructures, e.g., detected outages, BGP hijacks, DoS attacks, and other traffic anomalies, and meta-data to support analytics.<br>* Generate derivative data sets that reveal signals of connectivity disruptions from active and passive measurement methods.<br>* Experiment with which possible data sets are most amenable to live streaming to support HI-CUBE's near-real-time analytic capabilities.<br>* New data sets: logs of detected outages inferred from BGP, darknet traffic, and active measurements from Ark; and crowd-sourced measurements of networks vulnerable to IP source address spoofing |