Sisyphus or Sir Edmund? A Retrospective on Data Sharing

Erin Kenneally

Dept of Homeland Security Cyber Security Division
The (non-Oxford) Debate

Q1. Has data sharing improved over the past 3-5 yrs? By what measures?

[Pre: Yes=_____ No=_______ On Fence=_______]

Q2. Will Sir Edmund (improvements)/Sisyphus (deficiencies) advance in the next 1-5 yrs? Why?

[Post: Yes=_____ No=_____ On Fence=_______]
Motivating Questions

- What are the minimum components of sustainable data sharing?
- Are there good models of data sharing that get some element(s) right?
- Is a single model for data sharing desirable? Realistic?
- What should the public sector be doing to incentivize data sharing?
- Is sharing primarily a carrot/stick problem?
  - How to measure the ROI and effectiveness of data sharing?
  - Can data sharing (if done right) = competitive advantage?
  - Does data sharing need to be a legal and fiduciary responsibility?
• Coordinate, enhance and develop advanced data and information sharing tools, datasets, technologies, models, methodologies and infrastructure to strengthen the capabilities of national and international cyber risk R&D.

• These data sharing components are intended to be broadly available as national and international resources.

• To bridge the gap between producers of cyber-risk-relevant ground truth data, academic and industrial researchers, cyber security technology developers, and decision makers.

• In order to inform policy and analysis of cyber-risk and trust.
Whatcha Got For Me?

- Internet survey, hitlist, history and census datasets (USC)
- Internet outage datasets (USC)
- IDTK datasets (CAIDA)
- Packet captures (Natl Collegiate Defense Competition) (PCH)
- Malware passive DNS data daily feed (GaTech)
- Access to Internet Atlas (UW)
- US Long-haul Infrastructure Topology (UW)
- Two datasets carved from the DARPA 2009 scalable network monitoring program: attack stream from a target viewpoint and an attack stream from the viewpoint of the attacking botnet (CSU)
Name reflects an evolved implementation of the program's goals:

- **Marketplace** - A more open platform to connect and socialize data supply & demand;

- **Policy and Analysis** - Research infrastructure driven by and for real world issues; and

- **Cyber-risk & Trust** - Beyond just “defense” and “threats, an approach to sharing that views information as a critical infrastructure itself.
IMPACT STRATEGIC FOUNDATION:

- Data profile
- Central indexing

Metadata Discovery (FIND)

- Connect distributed suppliers & seekers of data & analytics

Data Matchmaking (GET)

- Get data to and from
- Optimize analysis

Social Networking (FEEDBACK LOOP)

Business Rules Broker

Tool Matchmaking (USE)

- Build collective knowledge
RESEARCH REQUIREMENT INPUTS

- Departmental Inputs
  - White House and NSS
  - State and Local
  - International Partners

- Critical Infrastructure Sectors (Private Sector)
- Interagency Collaboration

Departmental Inputs

Cyber Security Division
DHS S&T IMPACT Program
- How Can We Help?

Prioritization
Execution

Public Good
Value

Need

Defininition

>>

Apex Programs
Int'l Partners
State & Local
Private Sector

Network & System Security and Investigations
Int'l Cyber Infrastructure

Cyber Physical Systems

White House

State & Local

White House

Trustworthy Cyber Infrastructure

Applied Research
Advanced Development
Technology Transition

Defininition
>>

Prioritization
>>

Execution

Private Sector

Network & System Security and Investigations

State & Local

Apex Programs

Int'l Partners

Public Good
Your Turn to Share....

Data Marketplace Survey

15. What research questions or topics motivated your search for data? (check all that apply) *

- Peering relationships
- Internet economics
- Routing simulation
- Network security
- Complex networks
- System security
- Other
  
- Other
7. If you provide data to others, what type of data do you provide? *
Choose all that apply

- Address Space Allocation
- BGP Routing
- Blackhole Address Space
- DNS
- Infrastructure
- IDS and Firewall
- Internet Topology
- IP Packet Headers
- Performance & Quality Measurements
- Traffic Flow
- Sinkhole
- Synthetically Generated
- Unsolicited Bulk Email
- Not Applicable (I don’t provide data to others)
- Other
What is lacking with respect to your ability/desire to share data? *

<table>
<thead>
<tr>
<th>Option</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data exchange standards</td>
<td>★★★★★</td>
</tr>
<tr>
<td>Technical infrastructure (support &amp; maintenance)</td>
<td>★★★★★</td>
</tr>
<tr>
<td>Administration (dataset curation &amp; management)</td>
<td>★★★★★</td>
</tr>
<tr>
<td>Community coordination (e.g., feedback loop between data providers and seekers)</td>
<td>★★★★★</td>
</tr>
<tr>
<td>Data governance (standard rules and procedures for sharing data)</td>
<td>★★★★★</td>
</tr>
<tr>
<td>Legal assistance (knowledgeable advisement, accessible advisement)</td>
<td>★★★★★</td>
</tr>
<tr>
<td>Not Applicable (my ability/desire to share data is not lacking)</td>
<td>★★★★★</td>
</tr>
</tbody>
</table>
19. What are good examples of data sharing efforts that get some element(s) right? (e.g., SIE, Ren-ISAC, APWG, CsirtGadgets.org, SIE, CAIDA, CRAWDAD, Other) *

- SIE
- REN-ISAC
- APWG
- CSIRTGadgets.org
- CAIDA
- CRAWDAD
- Other
  - Other
  - Other
  - Other
20. In general, what is lacking with respect to a scalable (widespread, multiparty) ability/desire to share data? *
(check all that apply)

- Data exchange standards
- Data administration capabilities (curation, management)
- Technical infrastructure (support and maintenance)
- Community coordination
- Data governance
- Legal assistance

- Other

- Other
21. Do you think there is a need for a data exchange marketplace where data providers and seekers can go to list and acquire data and analytical tools? *

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>No Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
22. If you agree/strongly agree to the previous question, what are the elements of that marketplace that you believe are needed

<table>
<thead>
<tr>
<th>Element</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive, centralized metadata catalog of available datasets</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Social networking platform for data Providers, Consumers, and domain knowledge experts (technical, ethical, legal/policy)</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Standardized policies and procedures for finding and requesting datasets from federation of Providers</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Library of standardized templates and methodology for identifying and evaluating risks, intended utility, and applying disclosure controls on published dataset</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Centralized interface for standardized search and communications</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Centralized interface for requesting data from federated and/or trusted network of data providers</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Centralized interface to matchmaking analysis requests and responses</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Tools for analysis</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Tools for data management</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Feedback loop/reputation indicator for datasets available for exchange</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
</tbody>
</table>
23. Would you be willing to pay (money, data, other resource) for data in a marketplace exchange? *

- No
- Membership fees
- Subscription fees
- Individual dataset fees
- Sponsorship fees
- Data
- Other
IMPACT

v

R DATA

WE WANT YOU!
Help Us Improve Data Sharing: