

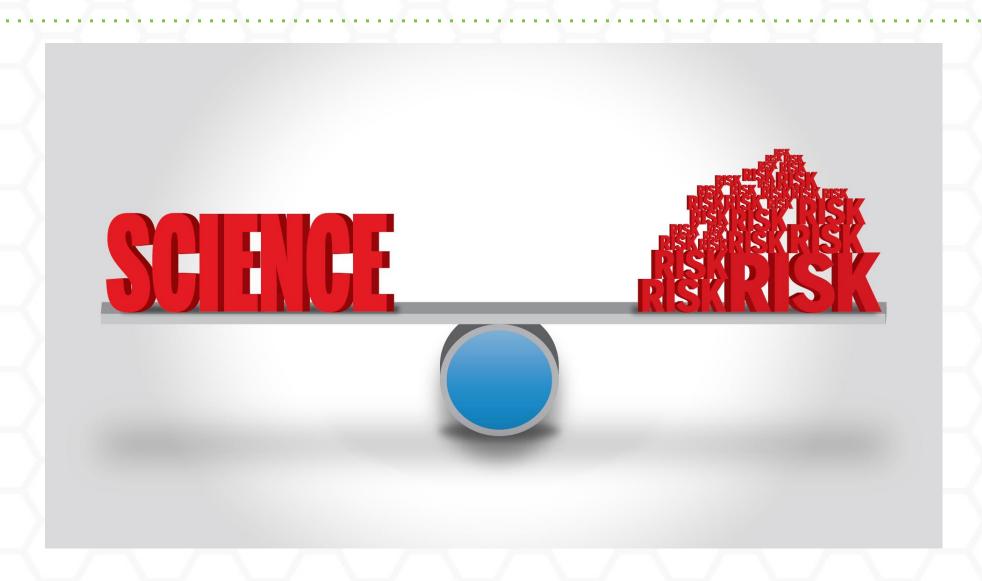
FAIR in an unfair world: Cybersecurity, data breaches, data integrity, and open science

Von Welch Director

International Symposium on Grids & Clouds 2019 (ISGC 2019)
April 4, 2019



What is the need for cybersecurity in open science and FAIR?



Trusted CI: The NSF Cybersecurity Center of Excellence

Our mission: to provide the NSF community a coherent understanding of cybersecurity's role in producing trustworthy science and the information and know-how required to achieve and maintain effective cybersecurity programs.

















https://trustedci.org/



Universities and research labs (IT and policies) CI,
Research
and
Education,
and
commercial
services

CI and open source software

Research and Education networks

NSF by the Numbers

~1,500 >\$1m

\$8B FY 2017 Budget Request

93% funds research, education and related activities





12,000 awards funded



2,000 NSF-funded institutions



362,000 people NSF supported



Fund research in all S&E disciplines



Fund STEM education & workforce





Trusted CI: Impacts

Trusted CI has impacted over 260 NSF projects since inception in 2012.

Members of more than 180 NSF projects have attended our NSF Cybersecurity Summit.

Members of more than 80 NSF projects have attended our monthly webinars.

We have provided more than 300 hours of training to the community.

We've had engagements with 41 projects, including nine NSF Large Facilities.



The Trusted CI Broader Impacts Project Report

June 28, 2018
For Public Distribution

Jeannette Dopheide¹, John Zage², Jim Basney³



https://hdl.handle.net/2022/22148

Regulated vs Open Science



Research with regulated data is guided by compliance

E.g. HIPAA, FISMA, NIST 800-171

Open science is not guided by compliance
E.g. Astronomy, climate, physics, geology
Most NSF science

A sizeable fraction or even majority of science at a University is open

If no medical school, probably majority.



This talk focuses on open science

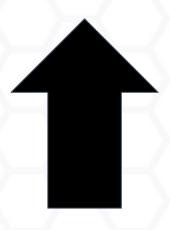
FAIR SCIENCE Cybersecurity



Trustworthy Productive Reproducible

Cybersecurity

Findable
Accessible
Interoperable
Re-usable



Cybersecurity

Findable, Accessible, Interoperable: Cybersecurity supports collaboration









Trustworthy Productive Reproducible



Re-usable

A Couple of Myths





Myth:

"I don't have confidential data, hence I don't need cybersecurity!"





The Role of Cybersecurity in Open Science and FAIR







Data Integrity

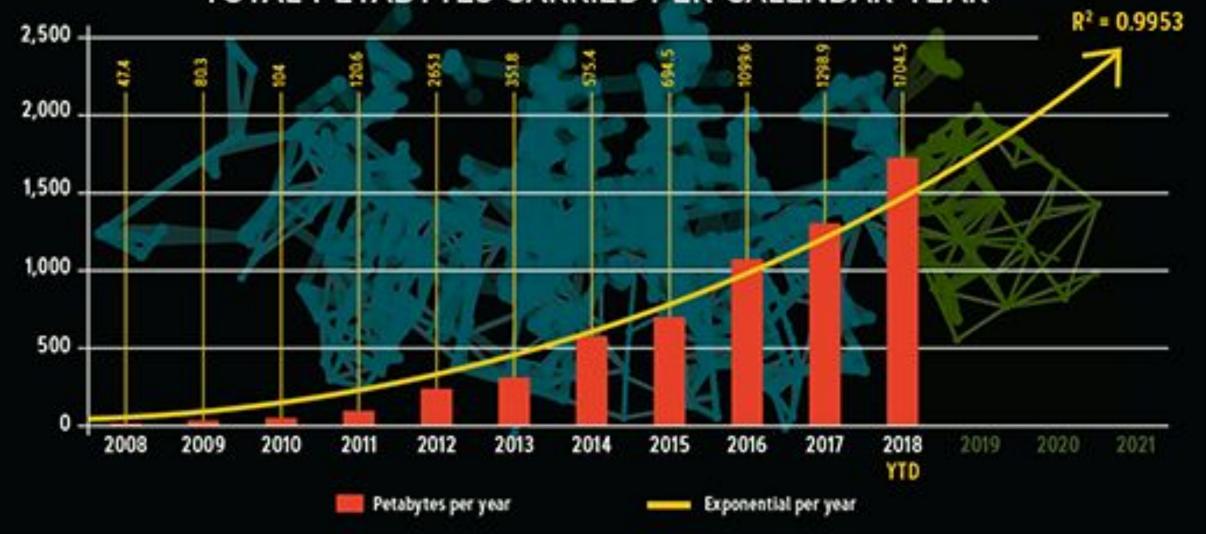
For open science, integrity of data is often most important aspect of cybersecurity.

Leads directly to trustworthiness and reproducibility of resulting science.





INTERNET2 NETWORK TOTAL PETABYTES CARRIED PER CALENDAR YEAR



CERN Study of Disk Errors

Examined Disk, Memory, RAID 5 errors.

"The error rates are at the 10-7 level, but with complicated patterns." (e.g., 80% of disk errors were 64k regions of corruption.)



Data integrity

Bernd Panzer-Steindel, CERN/IT Draft 1.3 8. April 2007

https://indico.cern.ch/event/13797/contributions/1362288/attachments/115080/163419/Data_integrity_v3.pdf

TCP Checksum Limits

"We conclude that the checksum will fail to detect errors for roughly 1 in 16 million to 10 billion packets."



When The CRC and TCP Checksum Disagree

Jonathan Stone Stanford Distributed Systems Group jonathan@dsg.stanford.edu Craig Partridge BBN Technologies craig@bbn.com

ABSTRACT

Traces of Internet packets from the past two years show that between 1 packet in 1,100 and 1 packet in 32,000 fails the TCP checksum, even on links where link-level CRCs should catch all but 1 in 4 billion errors. For certain situations, the rate of checksum failures can be even higher: in one hour-long test we observed a checksum failure of 1 packet in 400. We investigate why so many errors are observed, when link-level CRCs should catch nearly all of them.

We have collected nearly 500,000 packets which failed the TCP or UDP or IP checksum. This dataset shows the Internet has a wide variety of error sources which can not be detected by link-level checks. We describe analysis tools that have identified nearly 100 different error patterns. Categorizing packet errors, we can infer likely causes which explain roughly half the observed errors. The causes span the entire spectrum of a network stack, from memory errors to bugs in TCP.

After an analysis we conclude that the checksum will fail to detect errors for roughly 1 in 16 million to 10 billion packets. From our analysis of the cause of errors, we propose simple changes to several protocols which will decrease the rate of undetected error. Even so, the highly non-random distribution of errors strongly suggests some applications should employ application-level checksums or equivalents. We found this phenomenon of interest for two reasons. First, the error rate is disturbingly high. A naive calculation suggests that with a typical TCP segment size of a few hundred bytes, a file transfer of a million bytes (e.g., the size of a modest software down-load) might well have an undetected error. (We hasten to emphasize this calculation is naive. As we discuss later in the paper, a more realistic calculation requires an understanding of the types of errors.) Understanding why these errors occur could have a major impact on the reliability of Internet data transfers.

Second, there has been a long-running debate in the networking community about just how valuable the TCP (and UDP) checksum is. While practitioners have long argued on anecdotal evidence and personal experience that the checksum plays a vital role in preserving data integrity, few formal studies have been done. Studying these errors seemed a good chance to improve our understanding of the role of the checksum.

In this paper we report the results of two years of analysis, using traffic traces taken at a variety of points in the Internet. While we do not have a complete set of explanations (about half the errors continue to resist classification or identification) we can explain many of the errors and discuss their impact.

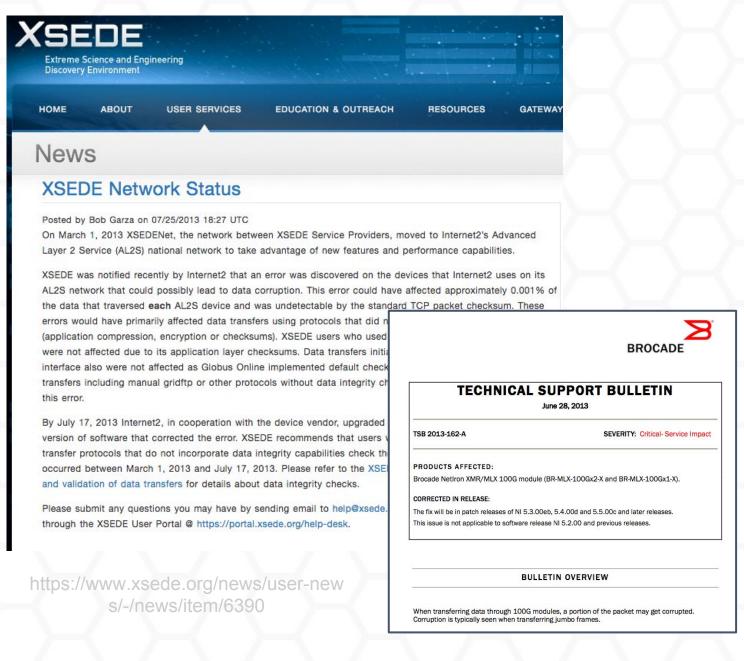
Network Corruption

Network router software inadvertently corrupts
TCP data and checksum!

XSEDE and Internet2 example from 2013.

Second similar case in 2017 example with FreeSurfer/Fsurf project.





Reproducibility

If your cyberinfrastructure isn't secure from unauthorized entities, reproducibility is at risk.

Need to manage tension between the need to patch vulnerabilities and the desire for stability to support reproducibility.



Threat of Unavailable Instruments

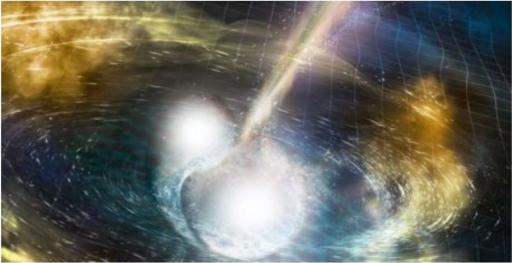
Cyber attack threatened WA astrophysicists' shot at gravitational waves, colliding neutron stars

NICOLAS PERPITCH

0:00

UPDATED TUE 17 OCT 2017, 6:44 PM AEDT





VIDEO [0:30] In a galaxy 130 million lights years away two neutron stars collide ABC NEWS

Astrophysicists at WA's Zadko telescope had just learned about the detection of a monumental deep space event involving two neutron stars colliding — which they had been hoping to find for years — when they came under sustained cyber attack.

At the critical and fleeting moment, they could not move their telescope to track the gigantic explosion 130 million light years away.

http://mobile.abc.net.au/news/2017-10-17/cyber-attack-almost-costs-team-look-at-colliding-neutron-stars/9055816?pfmredir=sm

Your Data Is Valuable to Criminals!



https://en.wikipedia.org/wiki/WannaCry_ransomware _attack

Rapid, Collaborative Projects

Research projects tend to be short-lived (3-5 years). They need to progress quickly. It's common for research collaborations to span universities and even countries.

Researchers want to define their teams, change those definitions and share access – all unrelated to institutional directories or human resources databases.

Interoperability is key.



Cyberinfrastructure != Enterprise IT

Secure Shell access to shared computers.

Uploading virtual machines, code, etc.

Science Gateways, Science DMZs

Distributed, high performance files systems, networks, etc.



Reputational Harm Will Erode Our Autonomy

COMMITTEE REPOSITORY

Calendar

Committees

Document Search

Hearing: Scholars or Spies: Foreign Plots Targeting America's Research and Development

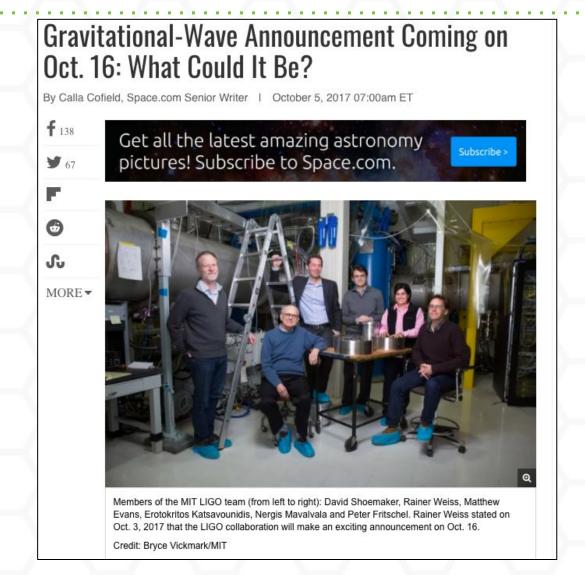
Subcommittee on Oversight (Committee on Science, Space, and Technology)

Wednesday, April 11, 2018 (10:00 AM)

2318 RHOB Washington, D.C.

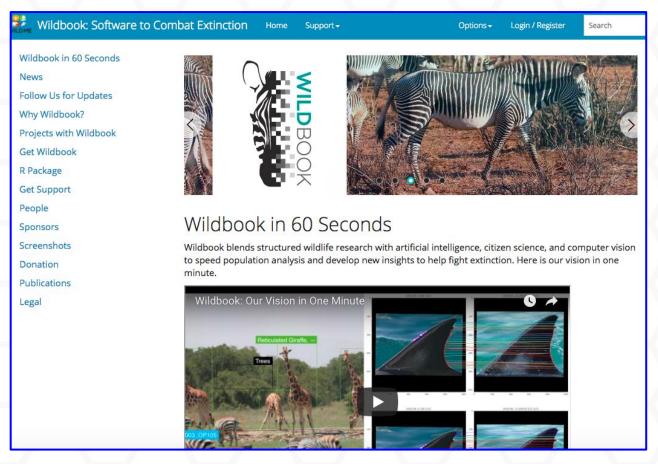


Confidentiality in Open Science: Pre-announcement/pre-publication



https://www.space.com/38367-gravitational-wave-announcement-coming.html

Confidentiality Driven by Ethical Concerns E.g. Endangered Species



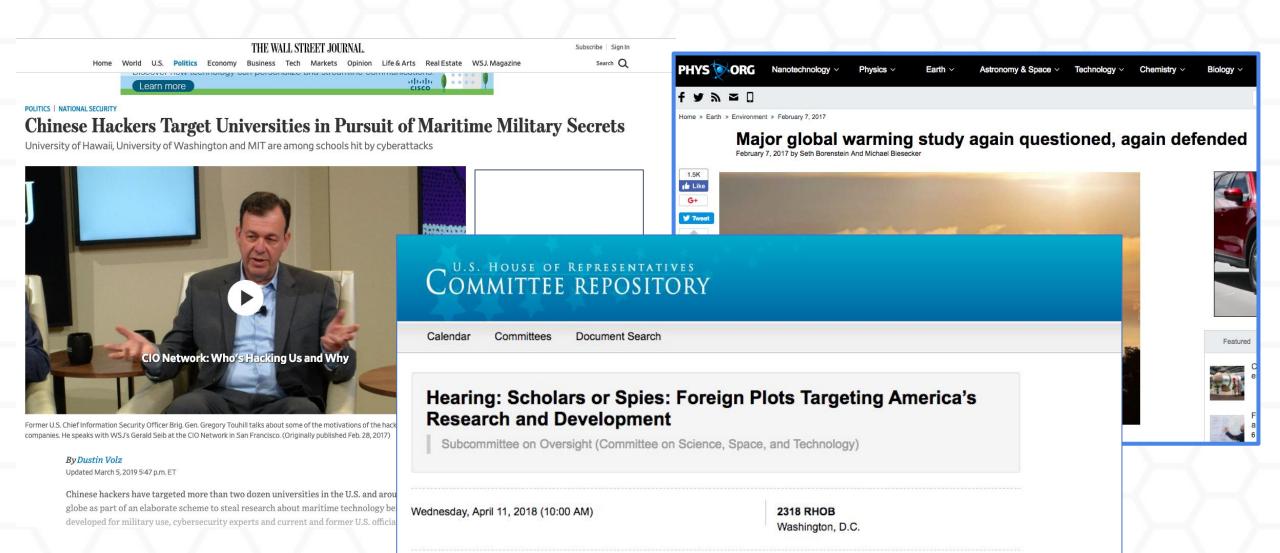
http://wildbook.org/



Is More Cybersecurity Better Cybersecurity for Science?



Increasing Risks and Political Pressure on U.S. Higher Education and Research



Appropriate for Science?

Abstract

[The errata update includes minor editorial changes to selected CUI security requirement additional references and definitions, and a new appendix that contains an expanded d about each CUI requirement.] The protection of Controlled Unclassified Information (CL resident in nonfederal systems and organizations is of paramount importance to federal and can directly impact the ability of the federal government to successfully conduct its missions and business operations. This publication provides federal agencies with a set recommended security requirements for protecting the confidentiality of CUI when such information is resident in nonfederal systems and organizations; when the nonfederal organization is not collecting or maintaining information on behalf of a federal agency or using or operating a system on behalf of an agency; and where there are no specific safeguarding requirements for protecting the confidentiality of CUI prescribed by the authorizing law, regulation, or governmentwide policy for the CUI category or subcategory listed in the CUI Registry. The security requirements apply to all components of nonfederal systems and organizations that process, store, or transmit CUI, or that provide security protection for such components. The requirements are intended for use by federal agencies in contractual vehicles or other agreements established between those agencies and nonfederal organizations.

"This publication provides federal agencies with a set of recommended security requirements for protecting the confidentiality of CUI when such information is resident in nonfederal systems and organizations;..."



Science need a Cybersecurity Framework that meets science's needs for productivity, trustworthiness, and reproducibility and is broadly accepted.



Trusted CI Path Forward



Create a Trusted CI Framework and Framework Implementation Guide for Open Science.

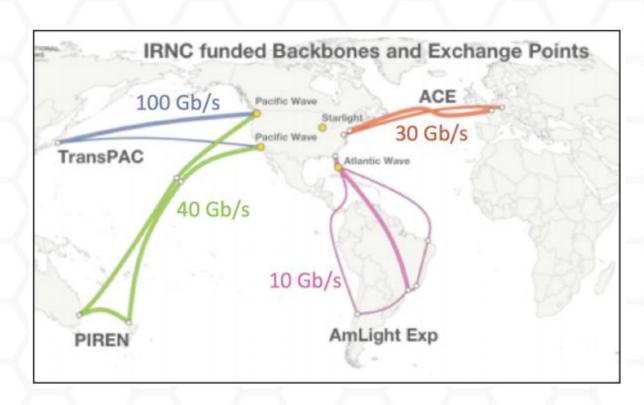
Builds on the Trusted CI Guide to Developing Cybersecurity Programs for NSF Science and Engineering Projects

https://trustedci.org/guide

A multi-year effort with early adopters and quick wins needed for success.



Framework Goal #1: Appropriate for Missions of Science



Gonzalez, A., Leigh, J., Peisert, S., Tierney, B., Balas, E., Radulovic, P., Schopf., J.M. "Big Data and Analysis of Data Transfers for International Research Networks using NetSage" IEEE Big Data Congress, SCF 2017

Integrity

Availability

Confidentiality

Framework Goal #2: Reasonable to Implement



- Medium-to-large projects
- Research centers
- Variety of science domains
- Limited cybersecurity workforce

Framework Goal #3: Broadly Accepted

Be accepted by

- Funding agencies,
- ClOs and ClSOs,
- Projects leads
- Auditors

as an acceptable cybersecurity program.



Universities
and
research
labs
(IT and
policies)

Research and Education, and commercial services

CI and open source software

Research and Education networks

Summary

Cybersecurity is key to FAIR and Trustworthy, Productive, Reproducible Science

We need a broadly acceptable cybersecurity framework appropriate for science.

Trusted CI is leading in developing such a framework.



Other Trusted CI Services

Cyberinfrastructure Vulnerabilities

Latest news on security vulnerabilities tailored for cyberinfrastructure community.

https://trustedci.org/vulnerabilities/

Specialized Information for Identity and Access Management, Science Gateways, Software Development

https://trustedci.org/iam/

https://trustedci.org/sgci/

https://trustedci.org/software-assurance/

https://trustedci.org/guide/

Large Facilities Security Team

Working group of security representatives from NSF Large Facilities.

https://trustedci.org/lfst/

Ask Us Anything

No question too big or too small.

info@trustedci.org

Follow Us

https://trustedci.org

https://blog.trustedci.org

@TrustedCI





Acknowledgments

Trusted CI is supported by the National Science Foundation under Grant ACI-1547272. The views expressed do not necessarily reflect the views of the National Science Foundation or any other organization.

Trusted CI activities are made possible thanks to the contributions of a multi-institutional team:

https://trustedci.org/who-we-are/



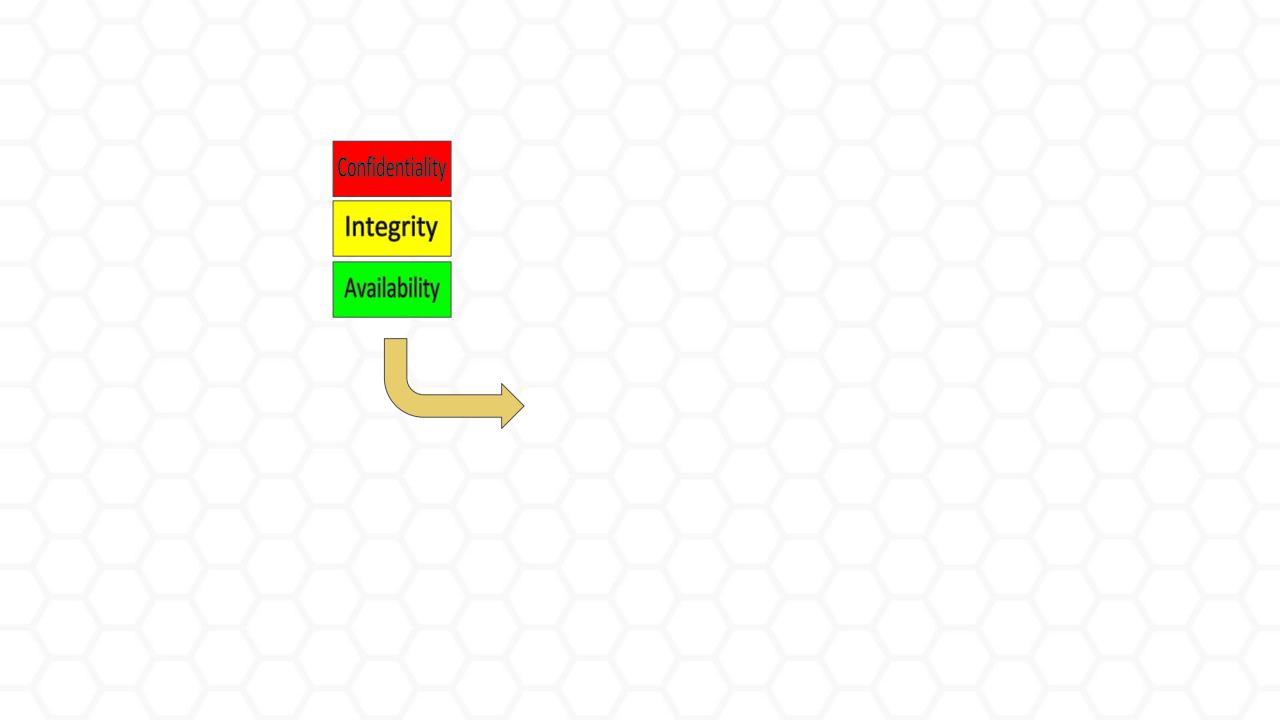






STOCK SLIDES FOLLOW



















Trusted CI: Impacts

Trusted CI has impacted over 190 NSF projects since inception in 2012.

More than 150 members of NSF projects attended our NSF Cybersecurity Summit.

Seventy NSF projects attended our monthly webinars.

We have provided more than 250 hours of training to the community.

Thirty-five engagements, including nine NSF Large Facilities.



The Trusted CI Broader Impacts Project Report

June 28, 2018
For Public Distribution

Jeannette Dopheide¹, John Zage², Jim Basney³



https://hdl.handle.net/2022/22148

Community-driven Guidance

Security Best Practices for Academic Cloud Service Providers

https://trustedci.org/cloud-service-provider-security-best-practices/

Operational Security

https://trustedci.org/guide

Identity Management Best Practices

https://trustedci.org/iam

Open Science Cyber Risk Profile

https://trustedci.org/oscrp/







Security Best Practices for Academic Cloud Service Providers

Version 1.0

http://hdl.handle.net/2022/22123



Engagements: One-on-one Collaborations

We take applications every six months.

Currently accepting applications for second half of 2019:

https://trustedci.org/application/

Deadline: April 3rd





Annual NSF Cybersecurity Summit

One day of training and workshops.

Agenda driven by call for participation.

Lessons learned and success from community.

Will be in San Diego in 2019.

https://trustedci.org/summit/





Trusted CI 5-year Vision and Strategic Plan

"A NSF cybersecurity ecosystem, formed of people, practical knowledge, processes, and cyberinfrastructure, that enables the NSF community to both manage cybersecurity risks and produce trustworthy science in support of NSF's vision of a nation that is the global leader in research and innovation."





The Trusted CI Vision for an NSF Cybersecurity Ecosystem

And Five-year Strategic Plan

2019-2023

Version 1

June 20th, 2018

Community Benchmarking

Some select results:

- Respondents' cybersecurity budgets vary widely.
- Respondents inconsistently establish cybersecurity officers.
- Residual risk acceptance is inconsistently practiced.



2017 NSF Community Cybersecurity Benchmarking Survey Report

> 8 June 2018 For Public Distribution

Scott Russell,1 Craig Jackson,2 Bob Cowles



https://hdl.handle.net/2022/22171

A Network of Cybersecurity Fellows

Fellows are liaisons between Trusted CI and communities.

Fellows receive training, travel support, and prioritized support.

Building on models from UK Software Sustainability Institute, ACI-REFs, Campus Champions.

Applications due: March 13, 2019

https://trustedci.org/fellows



Fellowship Programme

The Institute's Fellowship programme funds researchers in exchange for their expertise and advice.

The main goals of the Programme are gathering intelligence about research and software from all disciplines, encouraging Fellows to develop their interests in the area of software sustainability (especially in their areas of research) and aid them as ambassadors of good software practice in their domains. The programme also supports capacity building and policy development initiatives.

Each Fellow is allocated £3,000 to spend over



Campus Champions



Computational Science & Engineering makes the impossible possible; high performance computing makes the impossible practical Campus Champions Celebrate Ten Year Anniversary

Cybersecurity Transition to Practice (TTP)

Enabling researcher and practitioner collaboration to accelerate cybersecurity research to practice via

- matchmaking
- business model coaching
- workshops

https://trustedci.org/ttp



2019 Cybersecurity Transition to Practice (TTP) Workshop Wednesday, June 19th, 9am - 5pm. Chicago, IL

- Cybersecurity Topical Panels with Researchers and Practitioners
- Poster Session
- Thematic Co-creation breakouts for Research Transition to Practice

Request an invitation: https://trustedci.org/2019-ttp-workshop



Framework Pillars



Mission Alignment

• Information classification, asset inventory, external requirements

Governance

· Roles and responsibilities, policies, risk acceptance, program evaluation

Resources

People, budgets, services and tools

Controls

• Procedural, technical, administrative safeguards and countermeasures



Open Science Cyber Risk Profile (OSCRP)

OSCRP helps leads of science projects understand cybersecurity risks to their science and prepare for discussing those risks with their campus security office.

OSCRP was created by a team of computer security experts and scientists working together through a series of example use cases, which were then generalized to form the basis of the document.

OSCRP provides a mechanism for applying controls to mission-specific assets.

https://trustedci.org/oscrp/



OSCRP 2019 Planned Extensions

- 1. **Data integrity** issues in scientific computing, e.g., due to bit flips, are planned to be addressed.
- 2. **Data privacy and confidentiality (e.g., PII, proprietary technologies)** are planned to be explicitly addressed, including technical risk assessments.
- 3. Network-connected sensors and actuators ("cyber-physical systems") are planned to be examined in more depth.
- 4. Mitigations are planned to be included.
- 5. Cross references with the Trusted CI Framework will be added.



Trusted CI and Inclusivity

Cybersecurity requires diverse perspectives and cybersecurity community suffers from a lack of diversity.

Trusted CI works to address it through its workforce development, outreach, and community building efforts by explicitly seeking out and encouraging underrepresented groups to apply and striving for inclusive demographics.



2018 NSF Cybersecurity Summit Student Program



Trusted CI Partners























https://trustedci.org/partners

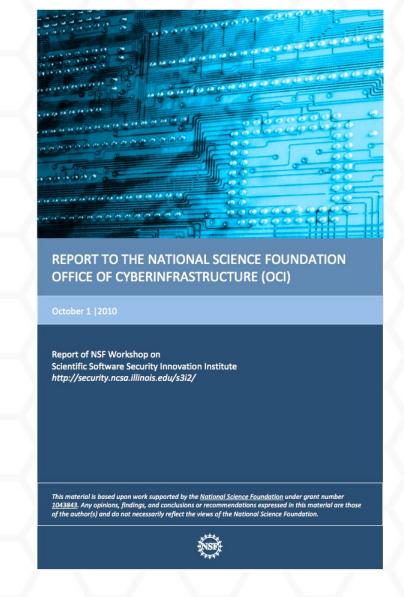
Extra Slides



We don't make the technology. We help you make sense of it.

Formed in 2012

Based on community call for leadership and guidance rather than technology







Harmonizing with SCI





Mission Alignment

Governance

Resources

Controls

SCI Areas

Participant Responsibilities

Data Protection

Operational Security

Incident Response

Traceability

https://wise-community.org/sci/





Research Security Operations Center
The second NSF-funded cybersecurity center serving the NSF
science community.

ResearchSOC complements Trusted CI



- Operational services and related training for NSF CI
- Community of Practice and Threat Intelligence Network
- Enabling Cybersecurity
 Research
- Outreach to Higher Ed Infosec regarding research CI



- Creating comprehensive cybersecurity programs
- Community building and leadership
- Training and best practices
- Tackling specific challenges of cybersecurity, software assurance, privacy, etc.











Operational cybersecurity services for research.

Building on existing services (OmniSOC, STINGAR) and expertise to bolster the NSF cybersecurity community's incident response capabilities.

Ramping up in 2019, initial clients in 2020, sustaining in 2021.

https://researchsoc.iu.edu/

NSF award 1840034