Welcome to the CCoE Webinar Series. Our topic today is the Open Science Cyber Risk Profile. Our host is Jeannette Dopheide.

The meeting will begin shortly. Participants are muted. You may type questions into the chat box during the presentation.

This meeting is being recorded.

The CTSC Webinar Series is supported by National Science Foundation grant #1547272.
The Open Science Cyber Risk Profile (OSCRP)

NSF CCoE Webinar
January 23, 2017

Rich LeDuc, Sean Peisert, Karen Stocks, Von Welch
A Scientist’s Perspective: the 3 stages of Cybersecurity

1. It’s open data.
   Nothing to worry about
2: Bad things happen
After OSCRP...

3. A foundation for preparedness
OSCRP Motivation and Goal

Our motivation is to help ensure the trustworthy nature of scientific computing by better understanding the project risks posed to science from cyberattacks. While much of Open Science has no confidentiality concerns, it is not free from risks related to computer-based attacks (cyber risks).

We are striving to go past the technological risks of hacked computers and understand when and if those risks translate to consequences to the science those computing resources support.

Hence, the primary goal of the OSCRP is to facilitate the understanding of how information technology risks relate to risks to the scientific mission of open science projects.
Bad Things Can Happen to Good Science

Untargeted attacks:

- Stepping stone - being used to attack others
- Ransomware - your data is valuable to you
- Because it has an IP address - worms, viruses, etc.

Targeted:

- Hackivism
- Political
- Scientific competition
Why OSCRP?

Our society is seeing an increasing level of sophistication of computer attacks that make scenarios which one would assume to have been confined to Hollywood real-world threats, including large-scale cybercrime and state-sponsored hacking.

- E.g. Stuxnet, the hack of Sony Pictures, the OPM breach of cleared individuals in the U.S., the attack on the Ukrainian power grid, the Yahoo breach of 1 billion user accounts.

It is also increasingly plausible to see open science as a target, as it includes both politically sensitive topics, valuable intellectual property, and areas with increasing privacy ramifications such as genomics and urban sensing.
Motivating Use Case:
The Consortium for Top Down Proteomics
The Consortium for Top Down Proteomics

- Mission: “To promote innovative research, collaboration and education accelerating the comprehensive analysis of intact proteins.”
- A collaborative effort between approximately 200 researchers worldwide
- [http://www.topdownproteomics.org/](http://www.topdownproteomics.org/)
Proteoform Repository

CTDP Proteoform Repository

NU IT

VM

Web Frontend

MySQL Backend

IU ScholarWorks

Data Archive
• NIH-funded (P41) research resource.

• The NRTDP provides a web portal using a Galaxy-frontend to NU IT Research Technology resources that allow authorized users access to Resource-derived HPC applications and workflows. The Resource also distributes several free data analysis applications.
NRTDP Infrastructure

NRTDP Cyberinfrastructure

NU Research Technology

VM

Galaxy web portal

Pulsar

NU Research Technology

Data Storage

HPC backend software

NU Quest Cluster

http://www.it.northwestern.edu/about/departments/cyb/
Challenges to OSCRCP Goal

In addition to commodity IT, science projects often rely on unusual, high performance IT and Internet-connected instruments for which risks are not broadly understood.

Open science has attributes that different from many other sectors - e.g. focus on integrity rather than confidentiality, cyber-physical component, open and distributed collaborative communities, reproducibility.

There is a knowledge and language gap between scientists and IT/Infosec pros that makes translating between science and cyber risks difficult.
**OSCRP Approach**

OSCRP’s approach is to provide an enumeration of common scientific assets and the IT risks associated with each.

- Scientific Assets == resources critical to science mission
- Common scientific assets is not a complete list. More later…

The desired usage is scientific and IT/infosec pro being able to sit down together, identify relevant science assets and their importance to the science mission, and then have a list of the cyber risks that matter. It is in effect, to be a bridge between the language of the scientist and the language of the infosec professional.
Asset-centric, Consequence-based

Actors, tactics, and vulnerabilities will change over time. Goal is for OSCRP to be useful for long time. Assume IT/infosec pros understand risks-of-the-day and focus on mapping science assets to those risks.

Hence:

- We started from science assets rather than IT threats.
- Focus on consequences to science mission rather than specific actors/tactics/vulnerabilities.
- Not a lot of detail on attacks and tactics.
The OSCRP Working Group

Leads:
Sean Peisert and Michael Dopheide, ESnet
Von Welch, Andrew Adams, and Susan Sons, NSF Cybersecurity Center of Excellence.

Core members:
RuthAnne Bevier (Caltech), Rich LeDuc (Northwestern), Pascal Meunier (Purdue/HUBzero), Stephen Schwab (USC Information Sciences Institute) and Karen Stocks (Scripps Institution of Oceanography).

Contributing members:
Ilkay Altintas (San Diego Supercomputer Center), James Cuff (Harvard), Reagan Moore (iRods), Warren Raquel (NCSA).
List of common science assets. Each linked to a diagram showing science concerns, consequences, and avenues of attack.

8. Common Open Science Assets

In this section we list common Open Science Assets and provide a diagram for each that provides a connection between the Concerns and Consequences from the perspective of the project mission and the Avenues of Attack that provide guidance to the cybersecurity lead in mitigating the Concerns regarding the Asset.

Data Assets

Data includes what we traditionally define as "data", i.e., raw data, derived data, as well as algorithms, protocols, configuration, accounting, personal or management data.

- **Public Data**: Published generated or collected Open Science data
- **Embargoed Data**: Open Science data that has not yet been publicly released
- **Internal data**: Data that is never intended to be published. This includes: generated or derived data (usually intermediate computational results), e-mails or notes
- **Documentation**: Manuals, wikis, blogs, etc. that facilitate the production, post-processing or management of Assets
- **Accounting Information**: Logs and databases recording the production, post-processing or management of Assets
- **For Approved Access Only**: Project/personnel data necessary for the Open Science mission

Facilities Assets

- **Facilities**: Physical storage, power & climate control used to house Assets
- **Staff computing & networking**: Institutional systems used by staff to access Assets, e.g., desktops, laptops, smartphones and the infrastructure they leverage (also mailing lists, calendars)
  - Note, the risk associated with this Asset is absorbed by the institution that owns them, not the project, but it is useful for the project to understand the greater risks involved

System and Hardware Assets

- **Networks**: Infrastructure used to access or transport data
- **File-store**: System to archive data
- **Front-end**: System to access archived data, usually web-based
- **Back-end**: System to search data, usually a database or source repository
- **User Portal**: System enabling users to initiate production or post-processing of data
- **Servers**: Systems used to access, store, produce and/or manipulate other Assets
- **Desktops**: Systems used to access, store, produce and/or manipulate other Assets
- **Laptops**: Systems used to access, store, produce and/or manipulate other Assets
- **Mobile devices**: Systems used to access other Assets (tablets, smartphones, smartwatches)
How to use the OSCRP?

1. Identify stakeholders
2. Work with stakeholders to create an inventory of important Assets.
3. For each mission critical science Asset, examine the Concerns, Consequences, and Avenues of Attack diagram associated with the Asset and note which Concerns and Consequences are relevant to the project, and the extent to which they are relevant.
4. For each relevant Concern note the vectors that could cause the Concern to be realized.
5. Work with IT professionals to agree on and implement agreed controls (protections against Concerns) or other means of mitigating risk.
   a. Specific threats and controls are out of scope of the OSCRP.
6. Repeat this assessment annually, or more frequently if the project’s Assets or risks have changed.
Usage Example: Embargoed Data Case Study

1) Identify the stakeholders: immediate project team, a collaborating research team, and the institutions that house the two telescopes.

2) Review OSCRP Asset Catalogue below and select relevant Assets:
   a) **Embargoed Data**: Due to the funding sources and collaborators involved, it’s important to the stakeholders that any potential discoveries made by this project have time to be vetted by appropriate experts and carefully released to the public. Specifically, the triangulated location of small and medium-sized near Earth objects should be kept secure for the private analysis of project scientists.
   
   b) **Staff Computing & Networking**: The workstations, laptops, and connectivity of project staff and any outside collaborators.
   
   c) **Telescopes**: Our telescopes have network-connected control system that allows for remote collaborators to adjust parameters. Both telescopes are considered similar, but separate Assets due to their placement at different institutions.
   
   d) **File-store**: The data from the telescope as well as preliminary results are stored on a central data storage system.
3) For each asset, look up relevant diagram in OSCRP.

E.g. Embargoed data:
http://trustedci.github.io/OSCRP/assets/Data/

Review concerns and consequences with stakeholders. Decide which are are most relevant to science mission.
Consider, Prioritize relevant Concerns and Consequences

Most concerned about lost data (red) with corrupted data (orange) being a close second. While we prefer our data to be embargoed, it would not be the end of the world if it happened to get exposed (green).

Next we are assuming for our example that it’s impossible to reproduce new data so the first Consequence from the reference diagram is removed, as well as the possibility of legal action.

Data exposure would result in another Consequence so we’ve added a custom one, “loss of control of result of publication,” in blue.
Avenues of Attack may expose other Relevant Assets

Science Asset: Embargoed Data

Concerns
- Inaccessible or Lost Data
- Not reproducible: lost science time and opportunities
- Inaccurate Science Results
- Reduced Reputation
- Legal Asset not regulated

Consequences
- Loss of control of result publication

Avenues of Attack
- Issues with sensor equipment (see Instruments Assets)
- Issues with data processing (see Software Assets)
- Issues with communication or storage (see Facilities Assets, or Systems and Hardware Assets)

Science Asset: Servers

Consequences
- Conditional on data type (see Data Assets)
- Device in inaccessible
- Device not performing as expected
- Device exposes embargoed or otherwise sensitive data during operation
- Equipment damaged
- Equipment lost, stolen or unreachable (see Facilities Assets)
- Network DoS affecting access (see Systems and Hardware Assets)
- Mis-configuration
- Unauthorized user
Step 5: We’re IT Security and We’re Here to Help

IT pros and Infosec staff can now focus on mitigating avenues of attack understanding their relevant to science Assets and mission.

Likewise, scientists will understand why the avenues of attack are meaningful to them.
Step 6: Lather, Rinse….. and Repeat.

Things will change. Assets will come and go. Mission will evolve.

Revisit annually to determine if Assets, Concerns, Consequences (or their priorities) have changed.

Revisit technical changes in the light of those adjustments.
OSCRP: A Living Document

New Assets will emerge. We will discover new consequences, concerns, and avenues of attack.

Hence OSCRP document is a living document, housed on GitHub (https://github.com/trustedci/OSCRP), distributed under a Attribution 4.0 International (CC BY 4.0) License, and designed to be easily contributed to.

Create an Issue to propose a change, or (ideally!) author your change and push it via GitHub Pull Request.

See: Section 14. Contributing to this Document.

Plan on “Releasing” 1.0 on February 10th.

Will revisit document quarterly to incorporate comments.
Conclusion

The goal of the OSCRP is to ensure the trustworthy nature of scientific computing by better understanding the project risks posed to science from cyberattacks.

By enumerating common scientific assets and mapping those to concerns, consequences and ultimately risks, we hope to enable a scientist and an infosec professional to discuss the scientific assets critical to a project and then translate the technical risks associated with those assets into risks to the project science mission and understand the severity of those risks.

We welcome community feedback and contributions to the OSCRP document.
More Information

The OSCRP document: https://trustedci.github.io/OSCRP/

For updates and the latest news, follow the CTSC blog: http://blog.trustedci.org/

Questions, or comments, please use a GitHub Issue: https://github.com/trustedci/OSCRP/issues
Or the CTSC-discuss email list: https://trustedci.org/ctsc-email-lists/
Thank you to the Working Group From the Organizers!

- RuthAnne Bevier (Caltech)
- Rich LeDuc (Northwestern)
- Pascal Meunier (Purdue HUBzero)
- Stephen Schwab (USC Information Sciences Institute)
- Karen Stocks (Scripps Institution of Oceanography).

And the contributing members: Ilkay Altintas (San Diego Supercomputer Center), James Cuff (Harvard), Reagan Moore (iRods), Warren Raquel (NCSA).

And presenters to the WG: Tanya Berger-Wolf, Matt Jones, Fred Luehring, and Alex Withers.
Acknowledgments

The NSF CCoE thanks the National Science Foundation (grant 1547272) for supporting their work.

Sean Peisert and Michael Dopheide’s work on this effort was supported by the Director, Office of Science, Office of Advanced Scientific Computing Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Karen Stocks’ work was supported by NSF OCE award 1327683 for the Science Support Office of the International Ocean Discovery Program.

This research was carried out in collaboration with the National Resource for Translational and Developmental Proteomics under Grant P41 GM108569 from the National Institute of General Medical Sciences, National Institutes of Health.

The working group thanks Tanya Berger-Wolf, Matt Jones, Fred Luehring, and Alex Withers for their presentations on the topic of science risks.

Any opinions, findings, conclusions, or recommendations contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of any of the sponsors of this work, employers of the authors, or any other organization.
Thank You!

Questions?

Please take our survey.
About the CTSC Webinar Series

To view presentations, join the discuss mailing list, or submit requests to present, visit:

http://trustedci.org/webinars

The next webinar is February 27th at 11am Eastern

Topic: Cybersecurity Program for Small Projects

Speakers: Craig Jackson, Susan Sons, and Robert Cowles

The CTSC Webinar Series is supported by National Science Foundation grant #1547272.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the NSF.
We thank the National Science Foundation (grant 1547272) for supporting our work.

The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the NSF.