**ITCR Training and Outreach Working Group Meeting**

**Thursday, November 08, 2018**

**4:00 - 5:00 PM Eastern Time**

**Agenda**

1. ITCR exhibition booth at ISMB2019 as outreach channel to feature software tools. All TOW participants, please express your interest [**HERE**](https://www.surveymonkey.com/r/G2X93Y3)
2. Review survey results about IT security. Please fill out this survey [here.](https://umichumhs.qualtrics.com/jfe/form/SV_06aTVWvpCd835bL)
3. Roundtable discussion: Continuation of discussion about IT security implications for software. What can the group do or can the NCI do to help address the challenges.

Link for brief survey about IT software security:

<https://umichumhs.qualtrics.com/jfe/form/SV_06aTVWvpCd835bL>

Please fill out before the end of the day on Wednesday November 7. It’s a short survey and should only take a few minutes. Much of the discussion will be focused on the results of the survey, so make sure your voice is heard!

Final Survey results (only 2 responses):

<https://drive.google.com/file/d/1EfFE3fIckeIcgo8fsEdQTbO637gg2D9X/view?usp=sharing>

|  |
| --- |
| **Webex Information:**  JOIN WEBEX MEETING  <https://cbiit.webex.com/cbiit/j.php?MTID=m90b62cf4b0ffc9b0c9e13b53425cc602>  Meeting number (access code): 733 094 302    JOIN BY PHONE  1-650-479-3207 Call-in toll number (US/Canada)    Global call-in numbers:  <https://cbiit.webex.com/cbiit/globalcallin.php?serviceType=MC&ED=574640482&tollFree=0> |

**Prior Meeting Notes:** [NCIP Hub](https://nciphub.org/groups/itcr/training_and_outreach_working_group_teleconference)

**Feedback about meeting or suggestions for future meetings:** [Anonymous Qualtrics Survey](https://umichumhs.qualtrics.com/jfe/form/SV_cIkVstbpgnbs4xT)

**Outreach Activities Since Last Meeting** (please fill in with new rows)

|  |  |  |
| --- | --- | --- |
| Name | Tool | Description of Activity |
| David Hanauer | EMERSE | Meeting/software demo with people at UCSF |
| Michael Reich | GenePattern Notebook | Lightning talk and poster, Cell Mapping Symposium, Cancer Cell Map Initiative, Sanford Burnham and UCSD |
| Rudi Pillich | NDEx | Outreach at CCMI Cell Mapping Symposium |
| Martin Morgan | Bioconductor | Nor’eastR Conference Presentation Semi-annual release! |
| Joseph Perl | TOPAS | Presentations at [Geant4-2018 | Third Geant4 International User Conference](http://geant4.in2p3.fr/2018/)  Also ran 3 new user trainings by video conference over last 2 weeks  Have now surpassed 500 users (34 countries) |
|  |  |  |

**Attendees, please sign in here:**

David Hanauer, U of Michigan

Michael Reich, UCSD

Mervi Heiskanen, NCI

Andrey Fedorov, BWH/HMS

Mary Goldman, UC Santa Cruz

Rudi Pillich, UCSD

Juli Klemm, NCI

Joseph Perl, SLAC/Stanford

Martin Morgan, Roswell Park Comprehensive Cancer Center

Amanda Bell, GW

**Minutes (everyone feel free to contribute):**

**ITCR booth @ ISMB 2019**:

* 4 responders will attend ISMB 2019
  + 3 participating attendees are interested in manning the booth
* 18 responders are not sure yet
  + 18 “not sure” responders are interested in manning the booth
* Pricing info coming soon
* If you haven’t taken the survey and wish to, please email me: [rudi@ucsd.edu](mailto:rudi@ucsd.edu)

Security Discussion

Fat fingered attacks: <https://blog.npmjs.org/post/163723642530/crossenv-malware-on-the-npm-registry>

from Andrey Fedorov to Everyone: 4:20 PM

might be useful as a summary of some of the tools developed by ITCR, to understand applicability of security considerations<https://docs.google.com/document/d/10k_ZgBaEBQesFRnAmYmEWPC9fQNrzSx0xI5oWXTcJcc/edit#heading=h.1vhdcjt3df96>

from Martin Morgan to Everyone: 4:29 PM

<https://blog.npmjs.org/post/163723642530/crossenv-malware-on-the-npm-registry>

from Andrey Fedorov to Everyone: 4:38 PM

here's one badge:<https://bestpractices.coreinfrastructure.org/en>

from Brion Sarachan to Everyone: 4:40 PM

An example software scanning company: <https://www.blackducksoftware.com/>

from Andrey Fedorov to Everyone: 4:41 PM

CII security requirements:<https://github.com/coreinfrastructure/best-practices-badge/blob/master/doc/criteria.md#security>

from Brigitte Raumann to Everyone: 4:46 PM

<https://trustedci.org/>

David Hanauer’s notes from the call:

Brian Zerkin from GE.

Often an open source component will include a lot, but they are only using a small percentage of functionality.

Martin from BioConductor:

They take packages contributed from authors, build them and make them available to users for download.

Think of the flow of the packages. Users makes them, transports them, build them and distribute them.

If you think about the integrity of the package (the user who builds it may be unknown to them). When these were distributed this was done via HTTP, you would get the package by typing a URL into the R session to download it. The package is a programming language than do lots of bad things. So there are opportunities to insert malicious things.

So there are even issues like "bioductor.org" instead of bioconductor.org.

So there were issues in the past.

So they have changed this to improve the process. For the distribution end, the primary way is now HTTPS. Now they are changing the way the packages are installed, so no URL is needed in the R session. It comes from a secure source and uses secures HTTPS for the packages.

The way the packages are transmitted to them come through a secure process.

They are sure the stuff they give them is what they give to the users.

But contributors are unvetted still.

Quite easy to write R code that does "nasty things". Still relying on the integrity of people distributing the software.

Q. Have there ever been attempts to use the system to generate attacks?

A. No.

For 3D slicer, platform is open source. Have a package contribution system. Contributors can maintain source code in their repository. It must be open source. The packages are then made available for download in the extension manager.

So this situation is applicable to 3D slicer platform.

Mary: Do users ask about this?

A: Some users have been concerned about this, and insecure HTTP distribution.

From the imaging analysis domain. Many applications are popular, downloaded over insecure channels.

Joseph Pearl from Topaz: Must think about the external packages that are being used.

Now, it is important to look if there are critical security updates. In the past don't want to disrupt users, but now maybe need to look for security vulnerabilities.

Brian from GE: In cases where software might be hosted in the cloud and let others upload their data. They become caretakers of someone else's asset. How do you protect someone else's data?

It is always the issue of priorities. For open source projects, too many unfunded tasks. Need to keep project alive and tradeoff of security vs more productive use of resources. People want new features, upgrades, handling submitted tickets.

Juli: From NCI point of view: Software must undergo quarterly vulnerability scans and software can be taken off online immediately if something is found. Maybe NCI is just an attractive target. Hosting software at the NCI is non-trivial if not exposed outside the firewall. Are other institutions developing a similar stance?

Speaker(?): He is sandboxed from patient data systems, etc. Lots of fuss about PHI security, but we still get patient information with the wrong e-mail address. Institutions must take responsibility for allowing research using software that doesn't meet the high security standards.

GE: Procedures for making code available. GE is also a target of many attacks. Must go through automated scans for 3rd party packages, has internal cybersecurity teams who will try to hack into any code that will be hosted to look for vulnerability. They can almost always hack into it, but question is how vulnerable it was.

Are there any companies to help with scanning/determining it "Secure"?

GE subscribed to "black duck" and can scan for 3rd party licenses.

NCI currently has no security expectation placed on ITCR-funded software. That is where it stands today. It is different for a government contract. Any software developed through contracts is different.

Bridget: NSF has an interesting model that they have taken advantage of. They fund cybersecurity institute. Part of the mandate is to work with other NSF-funded projects to improve the security of the software being produced. Every 6 months they have a call and you can apply for their services. They will do code reviews, help meet NIST standards, will do a whole number of security-related projects. This is only for grant-based work. Maybe NCI can do a similar model. Globus is a good example of this. It is part of national cyberinfrastructure.

Specifics: Several years ago they wrote a new component of their software that allowed for sharing files. They did a code review and looked at the security architecture, flow of certificates, authentication. They review it and the review becomes public. They do work with you before the report becomes public so you can address the issues. There were improvements that they were able to make. Process takes about 6 months long, and they take on about 6 projects every 6 months.

Proposal might be for designing software, and you need security consulting on what the architecture would look like.

It is so much easier to think about security from day one than to retrofit it for security later on.

Globus does a yearly endeavor for security reviews. They have a relationship with the University IT group/CIO. They are performing that service for free, since it is considered part of their mandate. It is not vulnerability scanning. It is mostly an operational review. Making sure they are operating software securely since they offer SAAS.

Research vs.clinical: Gray area: It is research data and it does have PHI. Not used for clinical decision making. So there still may be a liability and risk for breach.

Andrey: Imaging data may some day become completely identifiable since each image has a signature.

Juli: At NCI sequence data is considered possibly identifiable. This is an emerging area.