Slides: https://goo.gl/ZEyYDe

Mini Science DMZ (aka Mini-DMZ)

Steven Wallace
ssw@iu.edu
15-June-2018

Supported by the NSF via a CICI: Secure Data Architecture Award

Grad Students - are wonderful

Kaushik Srinivasan & Advait Marate, masters students at IU, have been great to work with. The future is good….except they’re leaving the project.

Inspiration

During our initial planning process, collecting use cases and user needs for IU's network network master plan, I was able to visit a number of research labs that contained scientific instruments. What we heard from those labs was the difficulty of attaching their instruments to the network due to security concerns.

The problem - Science Instruments are Insecure

Learned from Tracy Futhey To Consider adding Web access support

microscopes (crystallography, electron, optical, etc.) , flow cytometry, DNA sequencers, etc.

- Instruments are computer-based
- Most instruments are Windows computers
  - Can’t be patched
  - Can’t be upgraded
  - Are located randomly throughout campus
- Can be expensive to disinfect an instrument
- The instruments themselves can be very expensive, however, unlike HPC resources, may not be managed by cyber infrastructure specialist
- Data are born in instruments

The problem - Instruments Don’t support Provenance

There are exceptions, however these describe the norm:

- Metadata is the filename and/or the directory name
- There’s no check for data integrity (altered data is undetected)
- Data moves in and out of the science workflow via wetware
- No mechanism to support provence (i.e.,the data was created by what, when, where, and under the control of whom)

The problem - Instruments Don’t Make Good Test Points

- Some instruments can’t ping
- Nearly all instruments can’t be equipped with iperf
- Network impairments increase complexity of operating something that’s already unique
- Opportunity to leverage project to place PerfSONAR nodes at labs

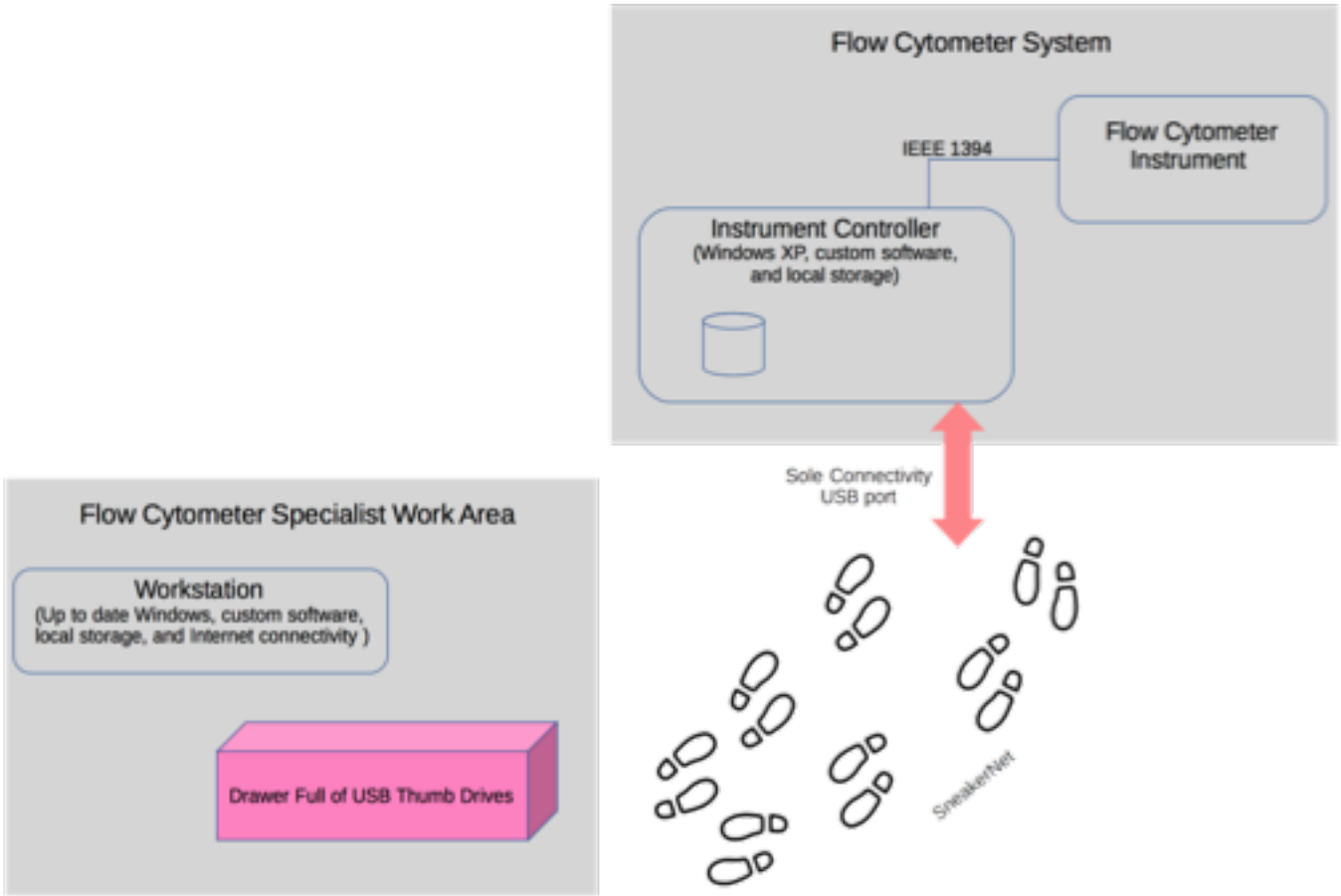
Typical of what we're finding: modest data size

Lots of bleach used here..

BSL2
Biologic Safety Level 2

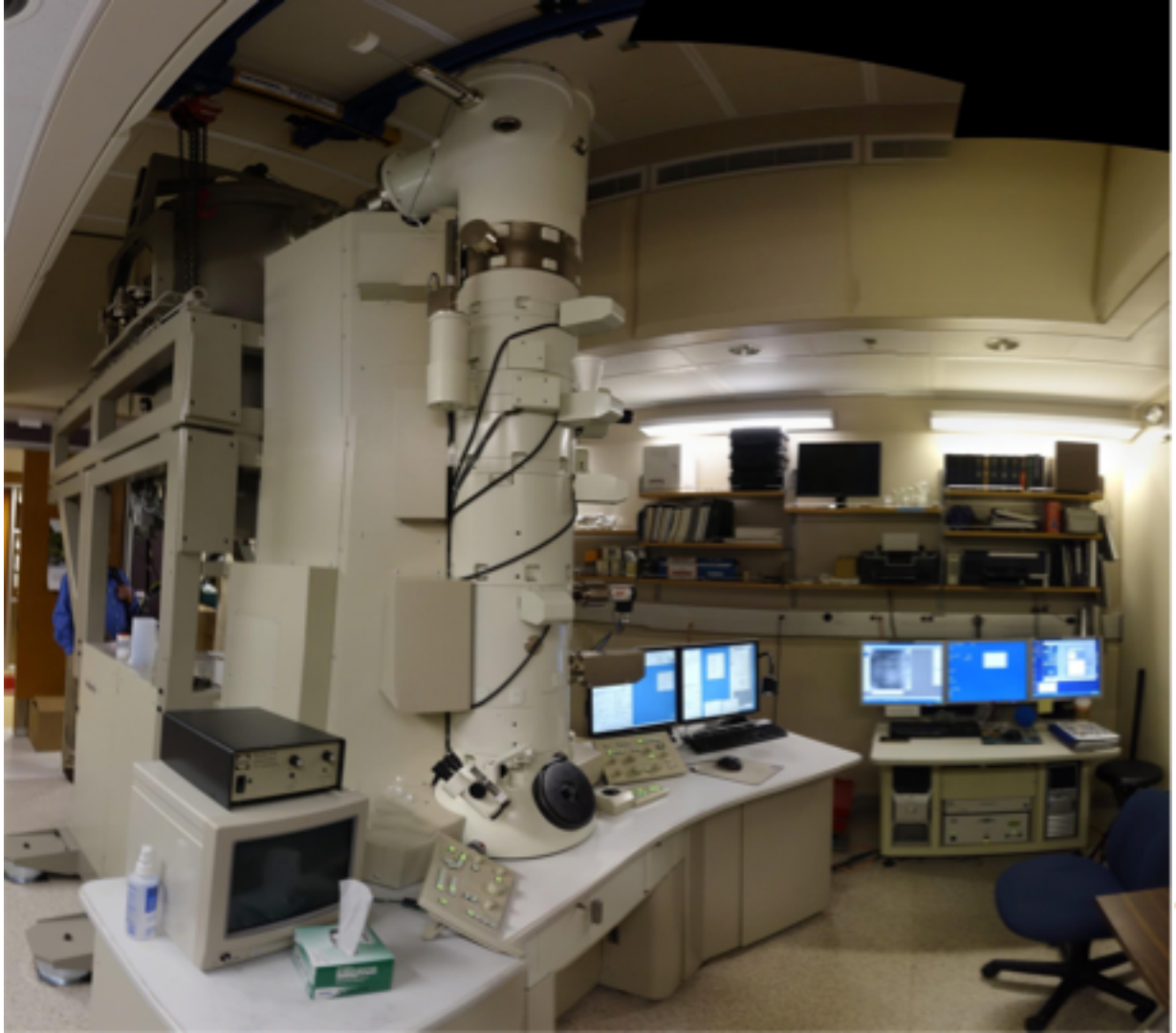
Department of Biology Flow Cytometry Core Facility

(current state 10-Aug-2016)



Electron Microscope

Upgraded sensor will generate 500Mb/s continuously



Capabilities

- Centralized configuration management
- Physical box (small form factor, perhaps ARM-based)
- Firewall & Intrusion Detection
- Data mover (facilitate data movement to science workflow)
- Data signer (digital sign data at creation)
- Network test point (partial perfSonar node)
- Protocol Proxy (e.g., DICOM)

Physical Box

- Small form factor (except for high-performance needs)
- Use case for affixing the Mini-DMZ to the instrument, and supergluing the cable connecting the instrument and the Mini-DMZ (no joke)
- Option for Power-over-Ethernet
- Headless, however LEDs status lights and/or small OLED display

Option for monitoring stuff in the lab? Secure Lab webcam?

Firewall & IDS

- Protect instrument and allow remote maintenance
- Seek to leverage existing solutions - Currently investigating pfSense (see: https://www.pfsense.org).
- Best outcome: Mini-DMZ become a supported product. Challenges include adding missing pieces to existing solution
- Also support encrypted tunnels & VPNs, potentially allowing remote instruments to appears local to campus network

Network Test Point

- Mini-DMZ will implement PerfSonar-TestPoint.
- Intend to create OAMP mesh to include campus PerfSONAR. OAMP likely limited to loss data given lack of stratum 0 time source and jitter of hardware such as a Rasp PI, however….
- Beaglebone may have little jitter, and may try DS3231-based clock along with NTP. Have others tried this?

Data Signer

- Cryptographically sign a blob of metadata that includes information about the instrument and the researcher, a secure hash of the data file(s), and a trusted timestamp
- In the future, a researcher can assert when, where, what instruments, keywords to aid future search, and when the data was created, as well as ensuring its [the data] integrity.
- Remarkably, this is a foreign concept to the researchers we've interviewed so far

[note: an IU security researcher suggested that researchers should sign and securely timestamp their hypothesis before they generate their data]

Data Signer

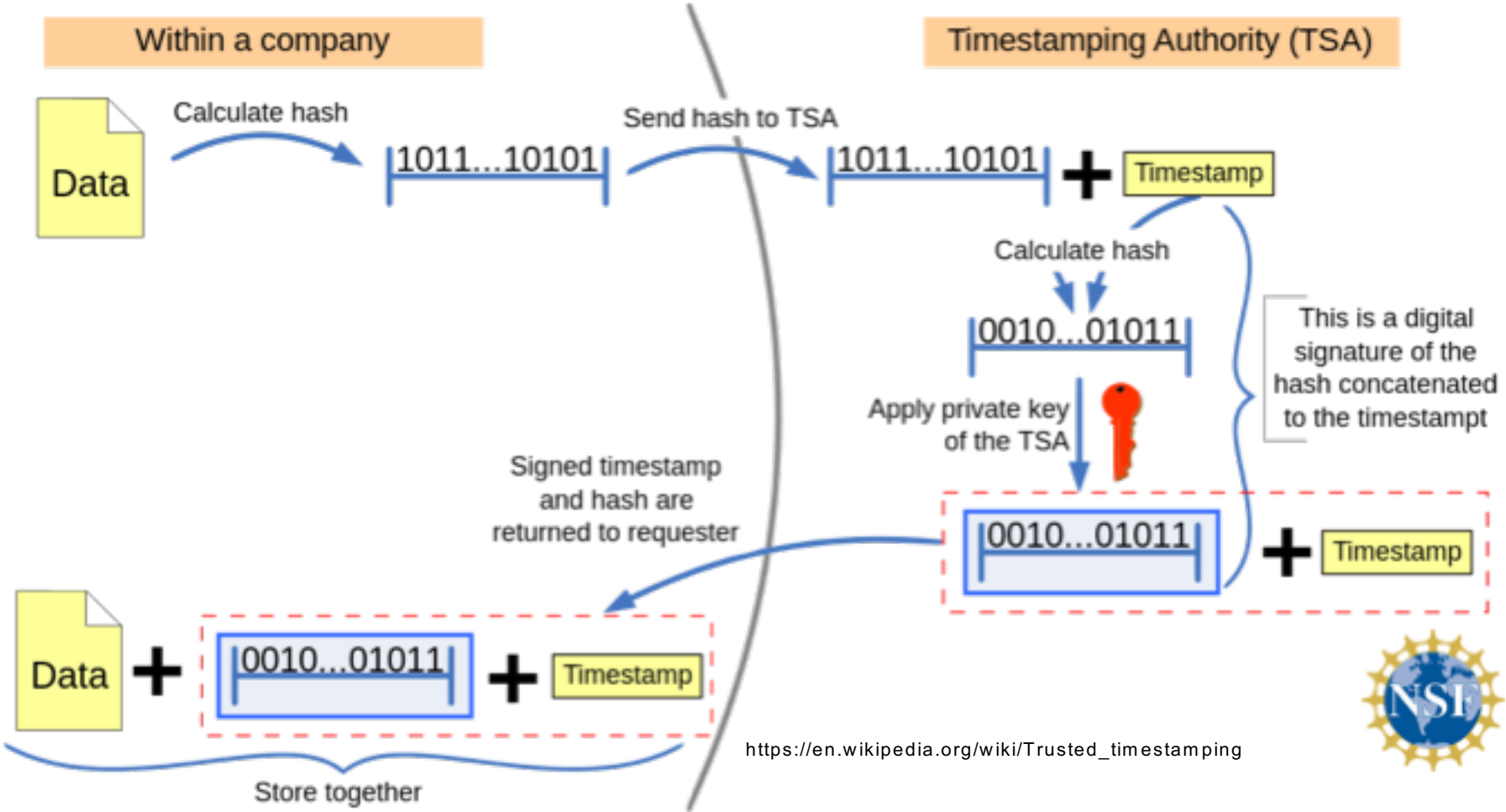
Check out: truetimestamp.org

http://truetimestamp.org/submit.php?auto=1&hash=**68b1a59a42f6f5713f960eced7abec70ab9f835fadc0dcd bad20b2a6f49bda7a**

Truetimestamp returns a text document that includes:

- The sha256 hash submitted above
- Time and Date
- PGP signature for verification
- Human readable instructions for verifying the PGP signature, even if truetimestamp.org disappears!

Trusted timestamping



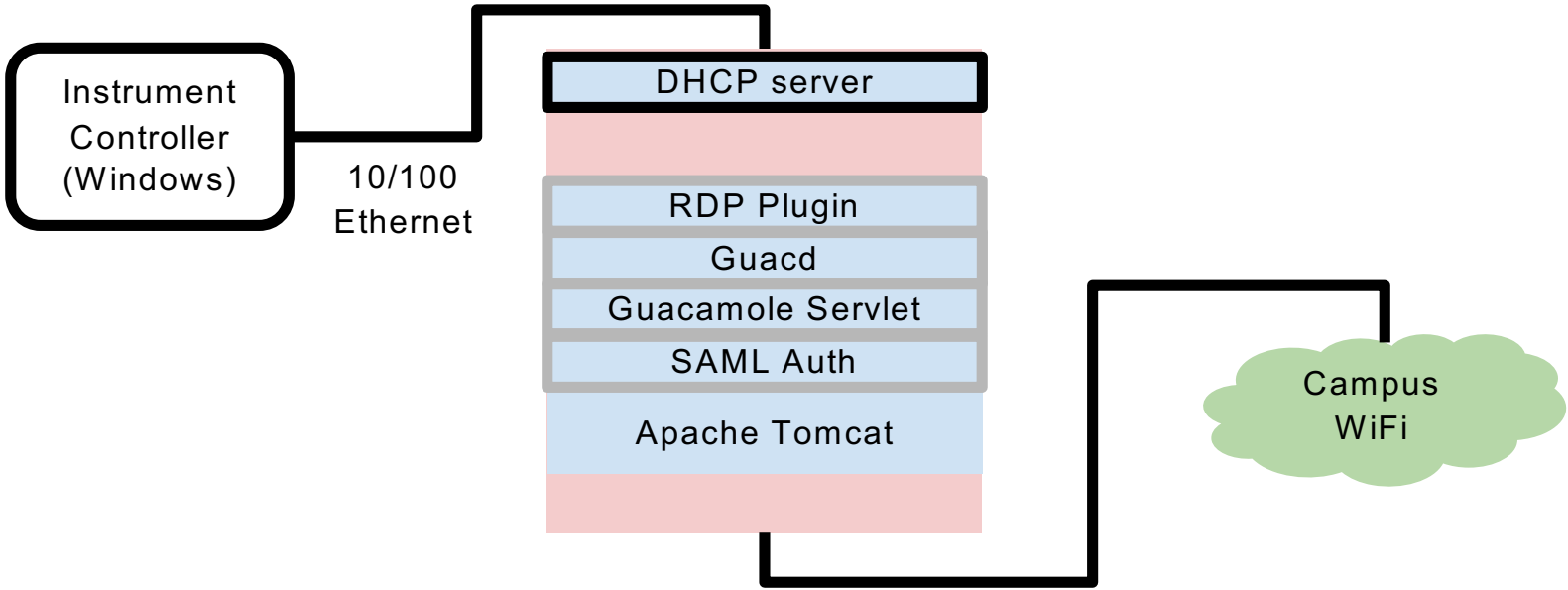
https://en.wikipedia.org/wiki/Trusted_timestamping

Side Question - Does our community desire its own TSA?

Data Mover

- Automate, to the extent possible, moving data created at/by the instrument into the science workflow.
- Data destination is arbitrary, often includes archive copy
- We have more use cases to review, however this appears to be challenging. Most data is moved via wetwear, executing a manual ad-hoc process, but a process that requires institutional memory.
- We intend to investigate existing work in this area, as well as attempting to find commonalities in a larger set of use cases.

Secure Remote Access



Protocol Proxy

For instruments that produce DICOM files (medical images), it may possible for the mini-DMZ to proxy the DICOM transfer protocol. Possibility an elegant mode for file moving to the science workflow.

Seeking other examples where a proxy may be a good approach.

Lots of Leveraging

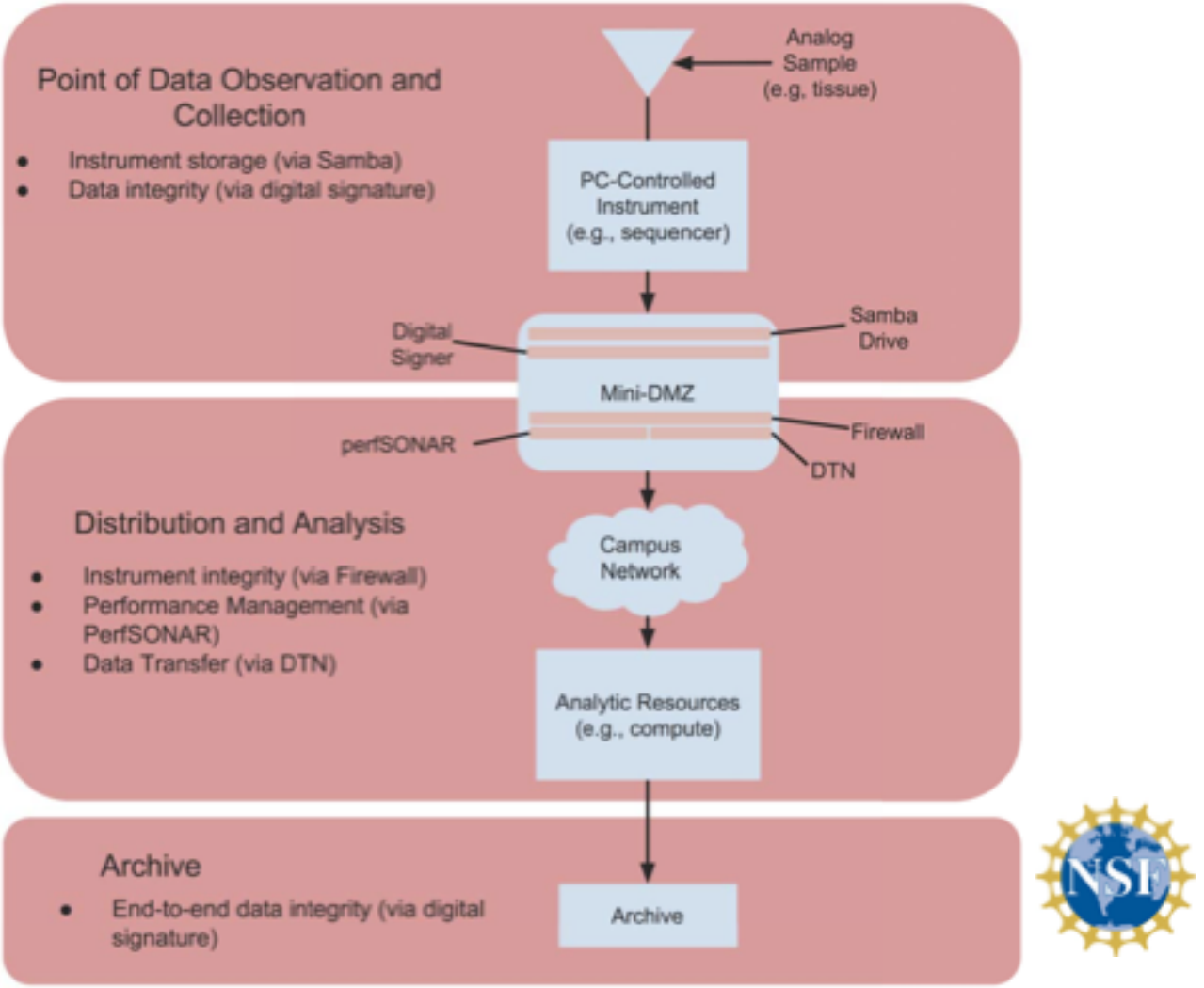
- PerfSONAR
- PerfSONAR mailing list
- Openssl
- Pfsense (or something similar)
- Globus Transfer API
- DCMTK (DICOM toolkit)
- Adafruit (precision clock, OLED display, etc.)
- PGP
- Ansible / Puppet
- Snort
- etc.

Long Tail Science….

- Continuing to wrap our head around metadata - science communities that share data understand its importance, other communities tend to see metadata as a nuisance. The provenance data is metadata, where does it fit?
- We anticipate “long tail science” will mature to normalize their data so that it becomes a community resource

Initial architectural vision

Remains roughly accurate



Project Status

- We have “working” code: https://github.com/kausrini/Mini-ScienceDMZ/tree/develop
  THIS LINK WILL CHANGE
- Core features in current beta:
  - Remote desktop access via client-less web-based service
  - PerfSONAR test node
  - CAS & SAML authentication
  - IPv4 and IPv6 access
- The project experienced false starts in its approach to authentication. Initial focus was on developing CAS plug-in for Guacamole. Lots of learning….we’ve now moved away from adding features to Guacamole.
- Local assumptions (i.e., sample size of 1) created technical debt.

Lessons Learned

Developing a turnkey security appliance is harder than I thought :-)

Larger sample sizes (diverse beta uses cases) are better.

Our security culture has changed since the proposal was submitted.

I’m convinced the MiniDMZ’s concept is sound, I’m concerned my project won’t translate into a sustaining benefit to researchers (sorry, just being honest).

But, all hope is not lost….

I have an idea: We [R&E community] should define the architecture and capabilities of a MiniDMZ, grounded in a diverse set of use cases...

Thanks!

Questions and Comments to: ssw@iu.edu