Mini Science DMZ (aka Mini-DMZ)

Steven Wallace
ssw@iu.edu

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

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

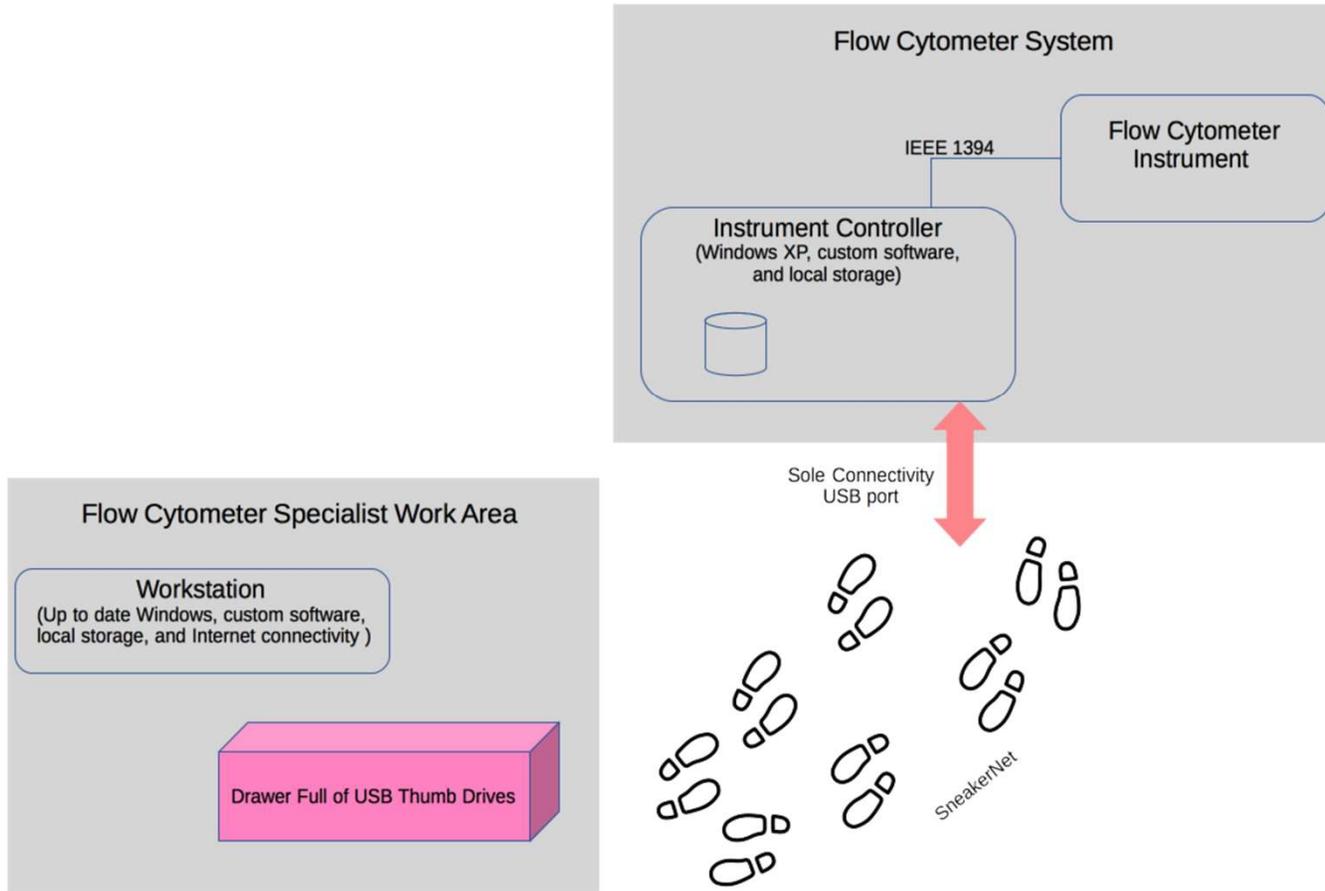Typical of what we're finding: modest data size

Lots of bleach used here..
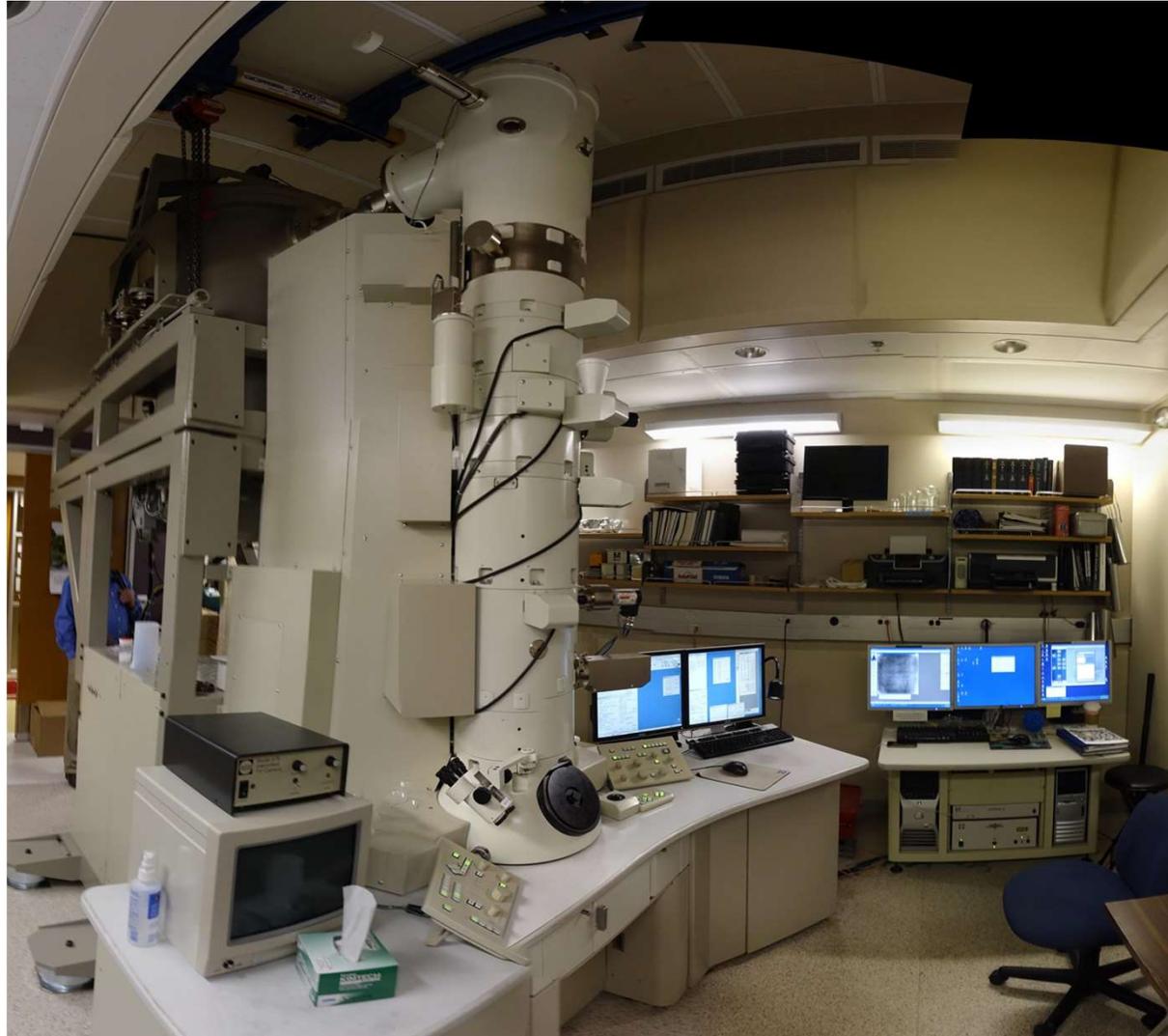
BSL2
Biologic Safety Level 2

Department of Biology Flow Cytometry Core Facility
(current state 10-Aug-2016)

Flow Cytometer System

IEEE 1394

Flow Cytometer Instrument

Instrument Controller
(Windows XP, custom software, and local storage)

Sole Connectivity
USB port

Flow Cytometer Specialist Work Area

Workstation
(Up to date Windows, custom software, local storage, and Internet connectivity)

Drawer Full of USB Thumb Drives

SneakerNet

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:
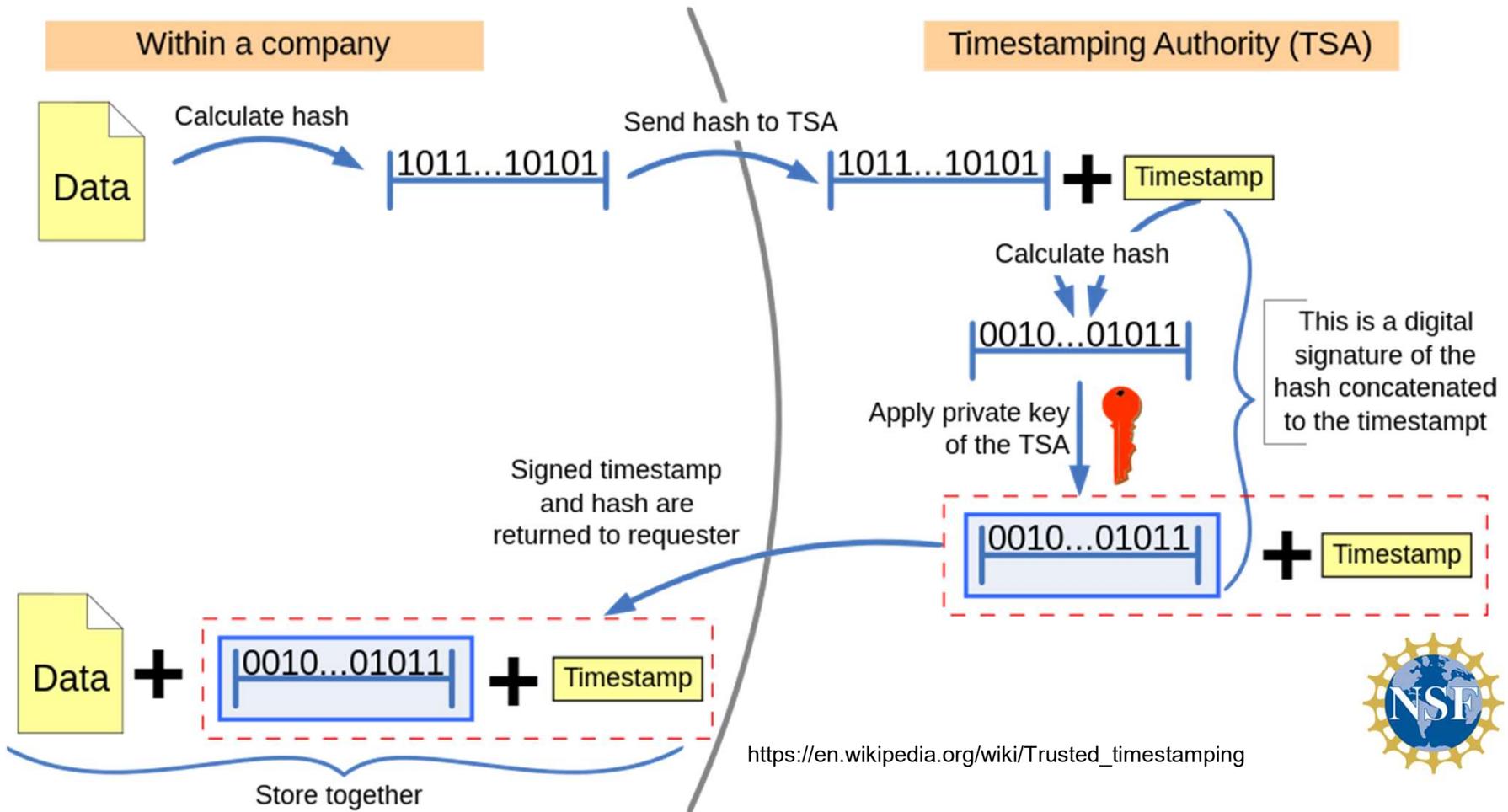
The sha256 hash submitted above

Time and Date

PGP signature for verification

Human readable instructions for verifying the PGP signature, even if
   truetimestamp.org disappears!

Trusted timestamping

Within a company

Data

Calculate hash → 1011...10101

Send hash to TSA →

Timestamping Authority (TSA)

1011...10101 **+** Timestamp

Calculate hash ↓↓

0010...01011

Apply private key of the TSA 🔑

This is a digital signature of the hash concatenated to the timestampt

0010...01011 **+** Timestamp

Signed timestamp and hash are returned to requester →

Data **+** 0010...01011 **+** Timestamp

Store together

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

NSF

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.

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

Project Status

Ongoing interviews to develop additional **detailed** use cases. This is key, and more challenging than anticipated.

Investigating hardware options, including pfsense appliances

Drafting architecture (not clear what parts are on the mini-DMZ itself vs. cloud or server based)

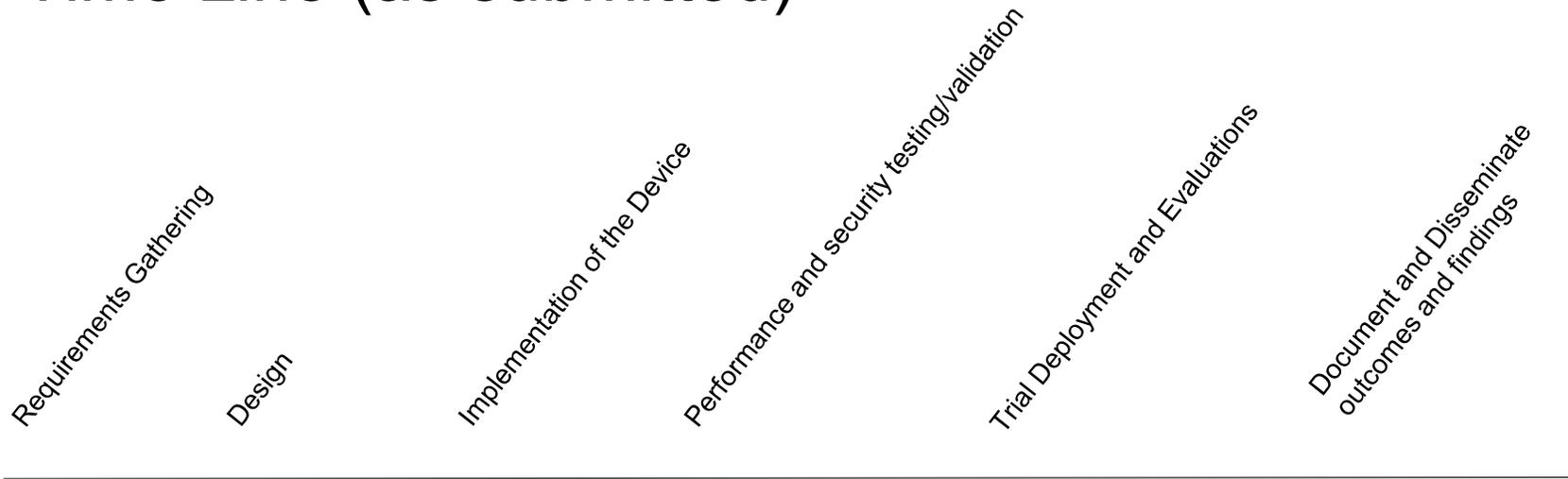Nearing the point of needing real programer contribution (not just me playing one)

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

Time Line (as submitted)

Requirements Gathering

Design

Implementation of the Device

Performance and security testing/validation

Trial Deployment and Evaluations

Document and Disseminate outcomes and findings

We're here
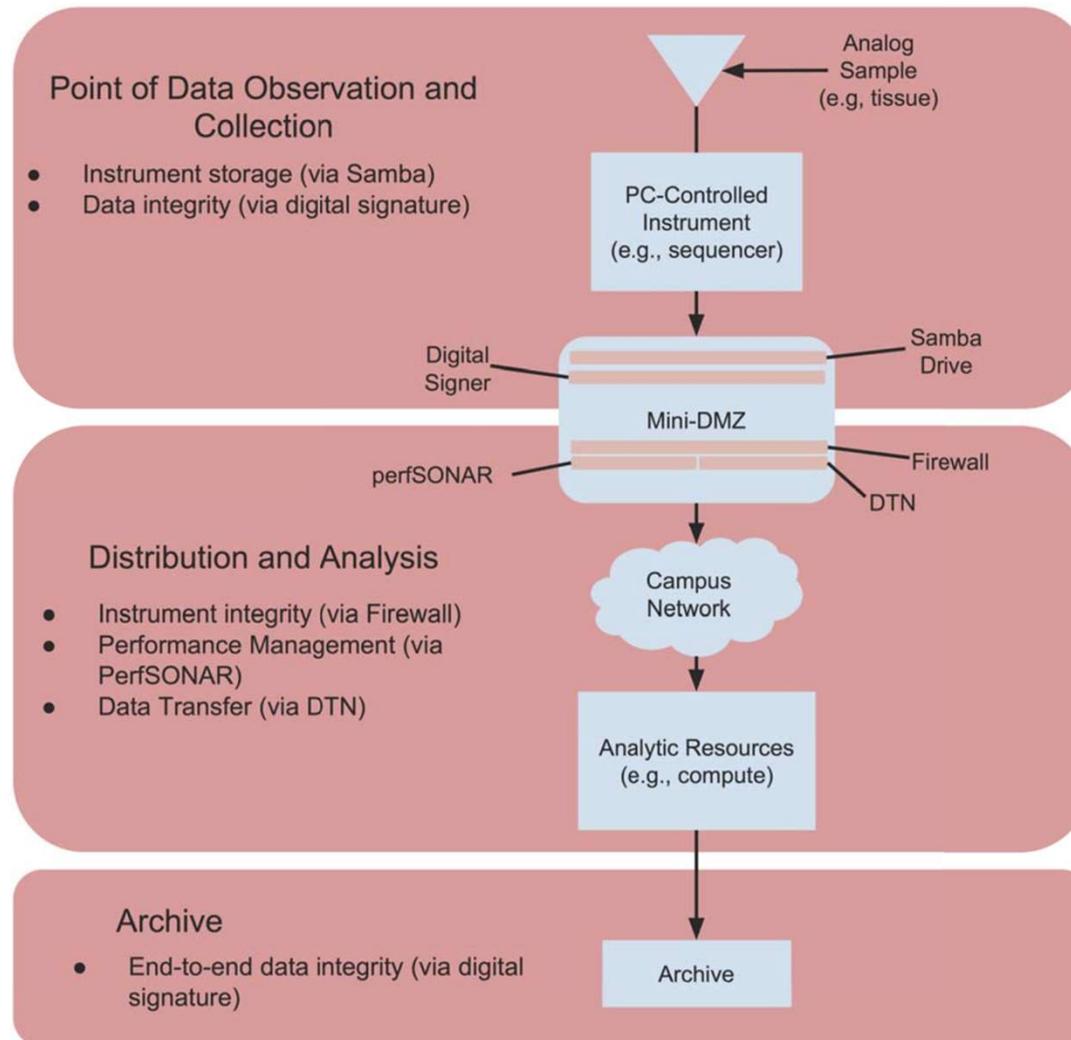
We should be here

3 years, but not to scale. Start of phase depicted

Initial architectural vision

Remains roughly accurate

Point of Data Observation and Collection
- Instrument storage (via Samba)
- Data integrity (via digital signature)

Analog Sample (e.g, tissue)

PC-Controlled Instrument (e.g., sequencer)

Digital Signer

Samba Drive

Mini-DMZ

perfSONAR

Firewall

DTN

Distribution and Analysis
- Instrument integrity (via Firewall)
- Performance Management (via PerfSONAR)
- Data Transfer (via DTN)

Campus Network

Analytic Resources (e.g., compute)

Archive
- End-to-end data integrity (via digital signature)

Archive

Questions Comments Discussion

Thanks!

Questions and Comments to: ssw@iu.edu