Governance, Risk and Compliance in the Globalized Scientific Community

September 2017

I'm Chris.

I'm an infrastructure geek.

I work for the BioTeam.

Twitter: @chris_dag

Content Warning

I am not an “expert”
… or a “thought leader”

I try to speak honestly about what I see, do and experience “on the ground” as an IT worker

My views are biased by the types of work I perform. Filter my words through your own expertise …

Goal this morning:

Talk candidly about risk, governance and compliance topics
… from our experiences *working directly with researchers & end-users*

Segue to Mark who can speak from the CSIO and Global Enterprise perspective

Topic: Humans

Human side of security, governance, risk management & compliance

Humans 01: Bridging the Language Gap

Domain specialists don't speak "my" language!

- **Language 'gap' becoming problematic**
  - 'Easy' when it was just Servers/Storage/VMs but now it's LANs, WANs, VPNs, Firewalls, IAM, Kerberos, AD integration minutiae etc. etc.
- **Even the word "risk" …**
  - Scientist: Risk = "Data Loss" or "Compute time vs. Publication Deadline"
- **BioTeam often called upon as Science/IT 'translators'**
  - [Informal] Joining meetings as SME and facilitating multi-party conversations
  - [Formal] Internal white papers written for specific audiences
  - [Formal] Post-incident ("*Why the firewall fell over when Research did X …*")

Humans 02: CapEx is easy. Headcount is hard

Not enough humans.

- **Easier to get infrastructure $ than human resources, thus …**
  - **Life science lacks the human capital necessary to properly engage in data classification, compliance, risk mitigation and incident response**
    - Result: Complex & expanding infrastructure run with “ops” focus
    - Result: Nobody doing strategic review or classification | systemic risk
  - **Example**: Pharma will expand storage capacity before hiring a human to properly curate, manage, classify and wrangle what they already have

Humans 03: Research vs Enterprise Resourcing

Critical teams are under-resourced and budgeted

- **Research often well budgeted w/ special appeal avenues**
  - Other less visible groups within are starved, shrunk or level-funded for years
- **Research can get ‘yes’ answers when seeking specialists**
  - Other groups get told “No … that is not your remit.”
- **Becoming a significant problem**
  - As peta-scale science diffuses out of the R&D organization and crosses LAN and WAN links to partners, clouds and collaborators, we increasingly need to rely on **Infosec**, **Security** and **Networking** groups that have been resource-starved for years.
  - Research Leadership MUST advocate for these groups !!!

Topic: On-premise Systems & Platforms

On-Premise 001: Identity Mgmt & Access Control

//

- **Active Directory is awesome (and I'm a Linux bigot) but …**
  - Inward-facing directories are no longer sufficient; more modern and flexible SSO and identity management, authorization and role-based access controls are needed
- **Need to manage identities & privileges in ways that span groups, organizations, entities and perimeter firewalls**
- **Resource Risk: Have you ever tried to get a meeting with the AD administrators at a global enterprise?**
  - There are usually ~2 people worldwide who truly understand the setup
  - And they don't take meeting requests from peons or nerds in R&D

On-Premise 02: POSIX insufficient

Owner|Group and Read|Write|Execute just does not cut it

- **Life science has been dealing with peta-scale data volumes for many years now**
  - Some of these files and data-sets are sensitive, proprietary, licensed only to named users or otherwise restricted in various ways
- **Filesystem access controls based on “group” and “owner” attributes are insufficiently flexible for the modern era**
- **Windows ACLs are more expressive but extending AD ACLs into Linux is painful and/or requires expensive proprietary software (ie Centrify)**

On-Premise 03: Folders & filenames as metadata

"Humans browsing directories" is no longer the core use-case

- **Issue #1: We still assume "humans browsing folders" is the dominant use case for scientific data at rest**
  - Not true when trillions of files are being produced and stored
- **Issue #2: We still use filenames, directory names and file/folder/sub-folder structures to implicitly supply metadata about a project and how it is organized**
  - Encoding organizational metadata via how folders are organized and named does not scale beyond Human-driven efforts.

On-Premise 04: Playing nice with Object Storage

The future of scientific data at rest is object based.

- **Mistake to create security, compliance and risk management methods that ASSUME the presence of a POSIX-style file and folder way of organizing data**
- **With trillions of files and petabyte+ volumes we must be planning for a future where directories do not exist & files have UUIDs and tags/metadata rather than descriptive filenames**

Security/Risk/Compliance should love object storage, 1/2

Object storage allows for use of custom/arbitrary ‘tags’ and metadata that can be indexed, searched and acted upon … huge win for automated compliance; These systems also have excellent ACL models and audit/access trail logging.

```
What instrument produced this data?
What funding source paid to produce this data?
What revision was the instrument/flowcell at?
Who is the primary PI or owner of this data? Secondary?
What protocol was used to prepare the sample?
Where did the sample come from?
Must this file be kept within certain geographic regions?
Where is the consent information?
Can this data be used to identify an individual?
What is the data retention classification for this file?
What is the security classification for this file?
Can this file be moved offsite?
etc. etc. etc.
…
```

Security/Risk/Compliance should love object storage, 2/2

A few features/capabilities of object storage …

- **Every event (put, get, delete, modify) is logged**
- **Unique credentials track every user, workflow & service account**
- **Supports modern Serverless/Lambda design patterns**
  - ***Automatic custom actions triggered upon any state change***
  - ***New file added? Trigger malware & tag compliance scan etc.***
- **Many different policy engines and integration hooks**
  - ***Security, Replication, Encryption, Sharing, etc.***

On-Premise 05: Perimeter Security Insufficient

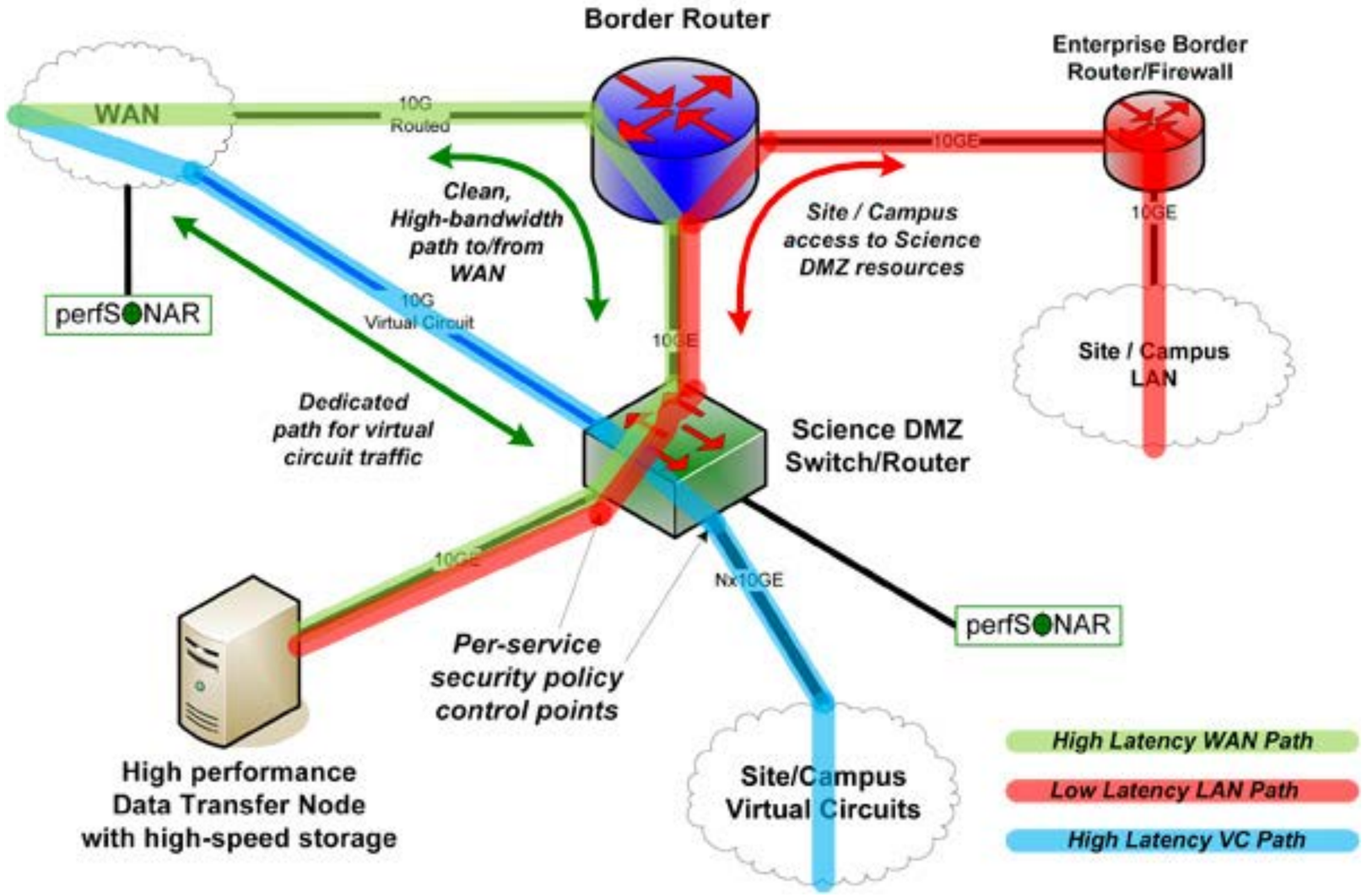
A firewall at the LAN/WAN edge is no longer enough

- **Our security architectures have not adopted for modern science-driven workloads. We still enforce a “hard-shell” via firewalls deployed at the LAN/WAN edge**
- **This does not fully address**
  - Isolation of research activities from “business traffic” on LAN
  - Diffusion of data, data movement and workflows between onsite & cloud
  - Increasing need for high-rate data flows in|out of organizations
  - Data exfiltration, malware, insider threats, ransomeware, etc.
  - Endpoint monitoring

On-Premise 06: Firewalls and “Elephant Flows”

Your firewall vendors may be lying to you

- **Firewall marketing is VERY misleading. That “10gig capable” firewall can’t actually handle a single 10gbps data xfer stream**
- **Enterprise security devices are not architected for ‘elephant flows’ and the new world of data intensive science**
- **Entrenched “checkbox culture” and platform monoculture means huge resistance to new techniques and methods when they are proposed to InfoSec & IT leadership**



https://fasterdata.es.net/science-dmz

Science DMZ as an example …

//

- **Emerging consensus around ‘Science DMZ’ design patterns**
- **These are real, not-vaporware and in production today**
- **High Speed & High Security are possible but …**
  - Implementation requires new methods, skills, products and techniques
  - … something that Enterprise Networking & InfoSec view with suspicion

On-Premise 07: Marketing vs Reality

Adventures in ‘secure analytics’ …

Marketing materials:

- Wicked awesome “data lake” & analytics platform
- Full encryption & RBAC
- Fully kerberized
- Full data governance, classification and security policy enforcement

Reality

- It works but …
  - Most security/compliance/governance features not enabled by default or covered in default install docs
  - Vendor consultants needed
  - Multiple rebuilds with configs provided by vendor, not docs
  - Required best Enterprise AD gurus and the “one person” who really understands kerberos
  - Multi-month implementation

On-Premise 07: Secured Analytics Project

A couple of good lessons we learned

- **This project required tying up senior experts and engineers from across the company for periods ranging from hours-days**
  - All the experts! Legal, AD, Kerberos, PKI/SSL, Oracle DBAs, InfoSec, Networking, Documentation & Process Leads etc. etc.
  - Some of whom rarely work with research or Research IT
- **Lesson #1: Complex cross-functional teams work**
- **Lesson #2: We found holes in our org chart & ops models when it comes to complex security, compliance & governance requirements**
- **Lesson #3: We can’t scale or sustain this level of activity without negatively impacting many other IT or portfolio projects**

Topic: IaaS Clouds & Discovery-oriented Research

A few slides about the c-word ..

Cloud 01: Life Science IaaS Cloud Drivers

Repeat After Me: This is not a cost saving play …

- **Cloud Driver #1 - Capability**
- **Cloud Driver #2 - Science/IT needs changing faster than datacenter**
- **Cloud Driver #3 - Collaboration & Data Exchange**
- **Cloud Driver #4 - Enterprise migration / transformation**

Amazon Web Services has opened case [redacted] on your behalf.

The details of the case are as follows:

Case ID: 222[redacted]
Subject: Your AWS account [redacted] is compromised
Severity: Low
Correspondence: Dear AWS Customer,

Your AWS Account is compromised! Please review the following notice and take immediate action to secure your account.

Your security is important to us. We have become aware that the AWS Access Key [redacted] (belonging to IAM user "buildmaul") along with the corresponding Secret Key is publicly available online at https://github.com/[redacted]

This poses a security risk to your account and other users, could lead to excessive charges from unauthorized activity or abuse, and violates the AWS Customer Agreement.

Please delete the exposed credentials from your AWS account by using the instructions below and take steps to prevent any new credentials from being published in this manner again. Unfortunately, deleting the keys from the public website and/or disabling them is NOT sufficient to secure your account.

To additionally protect your account from excessive charges, we have temporarily limited your ability to create some AWS resources. Please note that this does not make your account secure, it just partially limits the unauthorized usage for which you could be charged.

This is not an email you want to get ...

Amazon Web Services has opened case [redacted] on your behalf.

The details of the case are as follows:

Case ID: 222[redacted]
Subject: Your AWS account [redacted] is compromised
Severity: Low
Correspondence: Dear AWS Customer,

Your AWS Account is compromised! Please review the following notice and take immediate action to secure your account.

Your security is important to us. We have become aware that the AWS Access Key AKIAI[redacted] (belonging to IAM user "buildmgr") along with the corresponding Secret Key is publicly available online at https:\\/\\/github.com\\[redacted]

This poses a security risk to your account and other users, could lead to excessive charges from unauthorized activity or abuse, and violates the AWS Customer Agreement.

Please delete the exposed credentials from your AWS account by using the instructions below and take steps to prevent any new credentials from being published in this manner again. Unfortunately, deleting the keys from the public website and/or disabling them is NOT sufficient to secure your account.

To additionally protect your account from excessive charges, we have temporarily limited your ability to create some AWS resources. Please note that this does not make your account secure, it just partially limits the unauthorized usage for which you could be charged.

Cloud 02: Cloud Danger

When IT is no longer the ‘gatekeeper’ …

- **Major risks - I’ve seen these personally across multiple clients**
  - Some people are to senior to fire when they violate policy/rules
  - Data Loss / Public exposure of private data
  - Scientists care about efficiency/results and may not focus on security/risk
  - Keys/credentials/tokens leaking into internet or github repos
  - Accidental or intentional circumvention of geo-boundary rules
  - Nonfederated or ‘islands’ of disconnected identity and access control mgmt

Cloud 03: Major Cloud Advantages

It's worth it in the end …

- **Blunt Truth: IaaS cloud is more secure than your Organization**
  - Your policies and procedures don't reflect the real "ground truth"
  - *I've seen your datacenter loading dock door propped open for smokers*
- API-driven IaaS offers 100x more security, risk, monitoring, geo-fencing and audit-log features than traditionally found in on-premise resources
- If we use the cloud properly not only can we reduce risk and increase security we can often do this in ways that are far more powerful and far-reaching than what can be done on-premise with traditional method
- 100% API-driven means our monitoring, dashboarding, governance, compliance and policy-enforcement engines can be far more automated

Cloud 04: Cloud is changing all of our jobs

What to do when IT is no longer the gatekeeper?

- **Cloud is changing org charts and rewriting areas of responsibility**
- **IT is no longer a gatekeeper - the control plane is in the hands of the end-users and researchers. They decide what, where, how & how long**
- **The new role of IT is to architect and operate the environment and ‘safety guardrails’ within which the end-users operate**
  - … including all the ‘boring’ stuff that researchers don’t care about but is essential for the Organization
    - monitoring, reporting, risk mitigation, patching, security, event logging, compliance, incident response, etc.

Wrap up / Transition

Wrapping Up 1/2

My $.02 from being a practitioner on the ground …

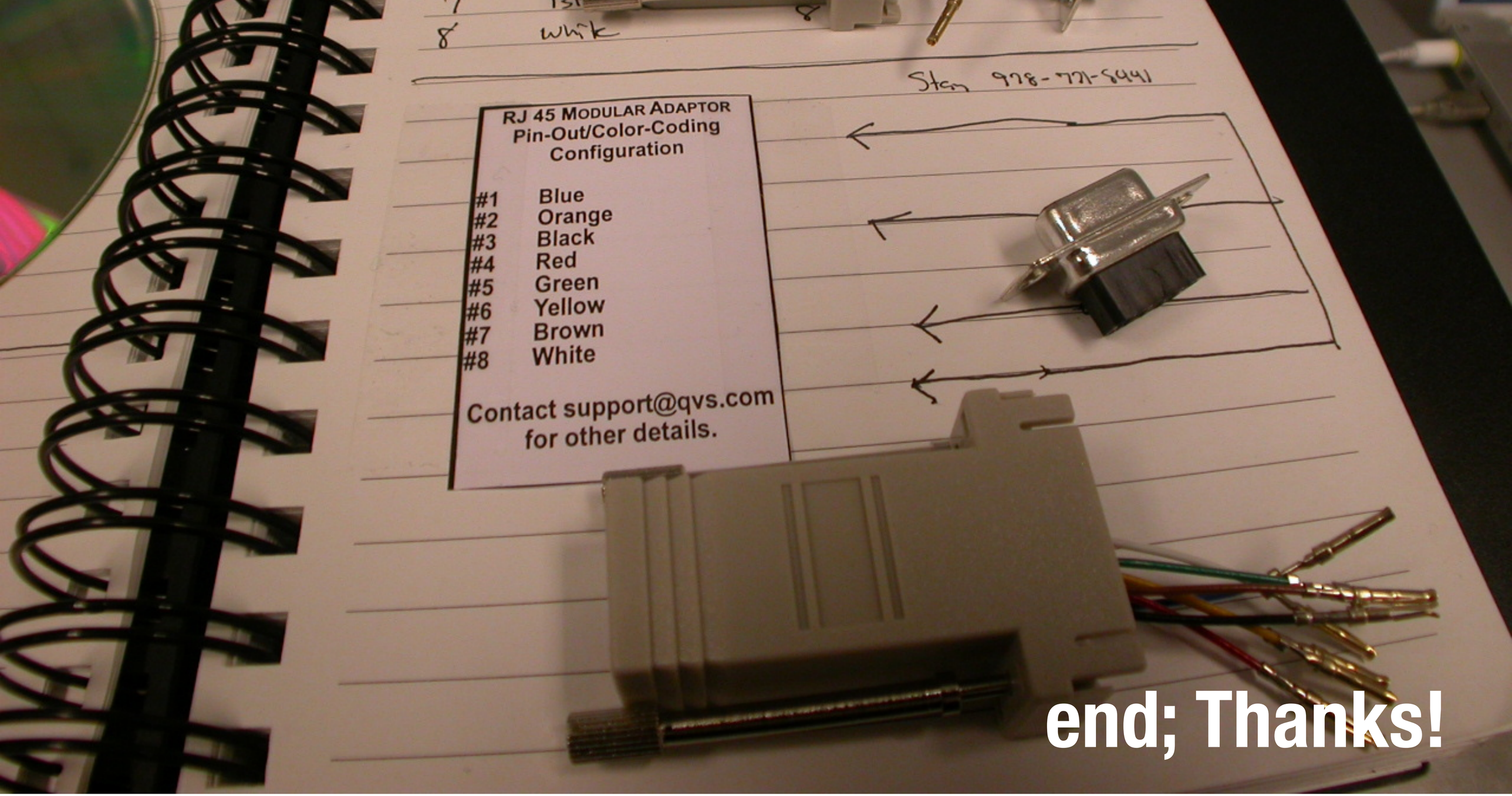
- **People:**
  - **“Language Barrier”**
  - **Still biased towards CapEx vs Humans**
  - **Security, Governance & Compliance greatly expands community of people research must engage with**
  - **These groups are horribly under-resourced; Research leadership must champion for these folks**
- **Infrastructure:**
  - **Still architecting hardware and SOPs that assume humans are primary file/data consumers**
  - **Still storing metadata via filenames & dir structures**
  - **POSIX is not the future**
  - **Perimeter-only security is a risk**
  - **Enterprise security kit often can’t handle ‘data intensive science’**

Wrapping Up 2/2

My $.02 from being a practitioner on the ground …

- Cloud Dangers
  - Significant risks abound when IT hands over infrastructure control capability to scientists, developers & end-users
- Role of IT will change
  - We will build the environment and manage the “safety net” while users control major elements

- Cloud Advantages
  - Better security and better security controls than on-premise if we are honest with ourselves
  - Security, compliance and risk mitigation features that we simply cannot replicate in-house
  - “API Everything” opens up new possibilities for automated compliance, policy & governance


RJ 45 Modular Adaptor
Pin-Out/Color-Coding
Configuration
#1 Blue
#2 Orange
#3 Black
#4 Red
#5 Green
#6 Yellow
#7 Brown
#8 White
Contact support@qvs.com
for other details.
Stan 978-771-8441

end; Thanks!