Modeling the Past and Future of Identity Management for Scientific Collaborations

Bob Cowles, Craig Jackson, Von Welch (PI)

ISGC 2016 – Taipei, Taiwan

15 March 2016

CENTER FOR APPLIED CYBERSECURITY RESEARCH

INDIANA UNIVERSITY

Pervasive Technology Institute

The XSIM Team

Bob Cowles – BrightLite Information Security, former CISO of SLAC.

Craig Jackson – CACR Policy Analyst, former practicing attorney.

Von Welch – CACR Director, long time distributed science security researcher.

Our Goal

Increase productivity of open DOE science through improved understanding of identity management (IdM) and relevant institutional risk.

Our Products

- Understanding of motivations and barriers to IdM responsibility sharing.
- Guidance for addressing barriers.
- Technical scheme for representing VO IdM sharing.

These products are freely available, we're happy to help with their application:

https://cacr.iu.edu/collab-idm/

The “Good Old Days”

Scientists were employees or students – physically co-located.

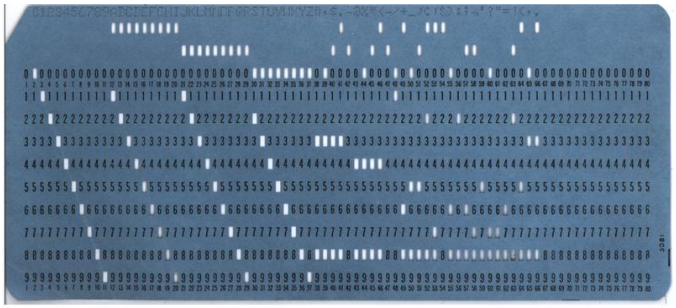
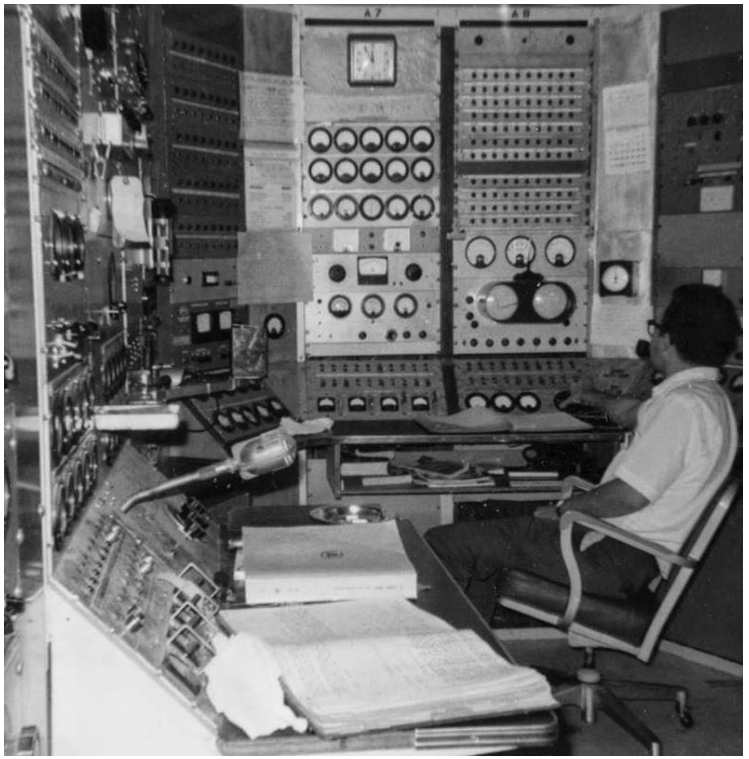


Image credit: Wikipedia



5

Then remote access…

Scientists start being remote from the computers.

But still affiliated with computing centers.

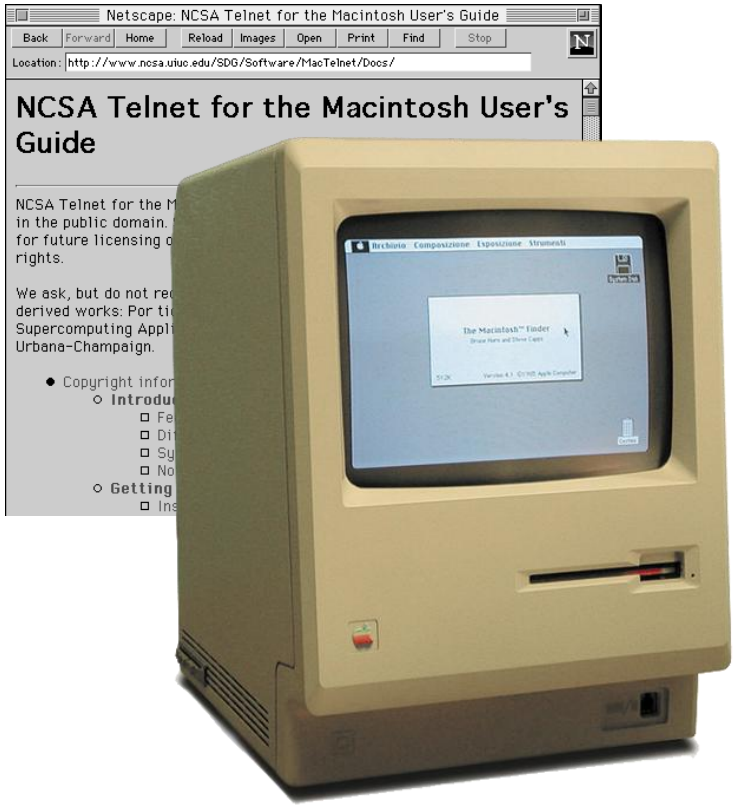


Image credit: All About Apple Museum
Creative Commons Attribution-Share Alike 2.5 Italy

Growth of the scientific collaboration

Number of scientists, institutions, resources.

Large, expensive, rare/unique instruments.

Increasing amounts of data.

Compact Muon Solenoid experiment at CERN's LHC

KBase PREDICTIVE BIOLOGY DOE Systems Biology Knowledgebase

LTER

ATLAS EXPERIMENT

FusionGRID www.fusiongrid.org

Some history of scale…

Date	Collaboration sizes	Data volume, archive technology
Late 1950's	2-3	Kilobits, notebooks
1960's	10-15	kB, punchcards
1970's	~35	MB, tape
1980's	~100	GB, tape, disk
1990's	700-800	TB, tape, disk
2010's	~3000	PB, tape, disk

Image credit: Ian Bird/CERN

Virtual Organization Identity Management

A number of approaches have been tried:

VOMS, Glide-ins, Science gateways, COManage, Community/group accounts, etc.

We have 15 years of applied experimentation in virtual organization (VO) IdM.

Twenty+ Interviews

VOs

- Atlas
- BaBar
- Belle-II
- CMS
- Darkside
- Engage
- Earth System Grid
- Fermi Space Telescope
- LIGO
- LSST/DESC

Resource Providers

- Atlas Great Lakes T2
- FermiGrid
- GRIF
- U. Nebraska (CMS)
- LCLS
- RAL
- GRIF/LAL
- LLNL
- NERSC
- Blue Waters

IdM is Critical to Science

- Control of unique instruments
- Ability to QA data
- Access to pre-publication data
- Membership and structure of collaboration
- Names on papers
- Etc.

<u>However, scientists don't use IdM's nomenclature!</u>

Data-centric Model for IdM Sharing

Offers a common language and graphical representation to complex IdM requirements/implementation

Goal: Facilitate communication between scientists, IdM, CSO.

Functionality

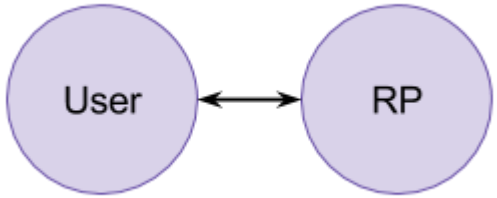
authentication
authorization
allocation/scheduling
accounting
auditing
user support
incident response

Model IdM Data

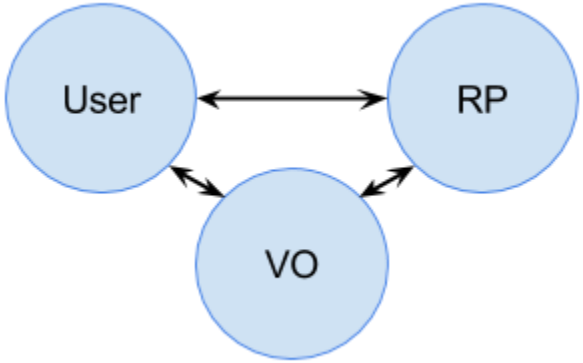
(1)User identifier
(2)User contact info
(3)VO membership/role

Transitive Trust

Classically RPs produced and consumed all IdM data.



Brokered trust relationships entail VOs & TTPs generating user data, to be consumed by RPs.



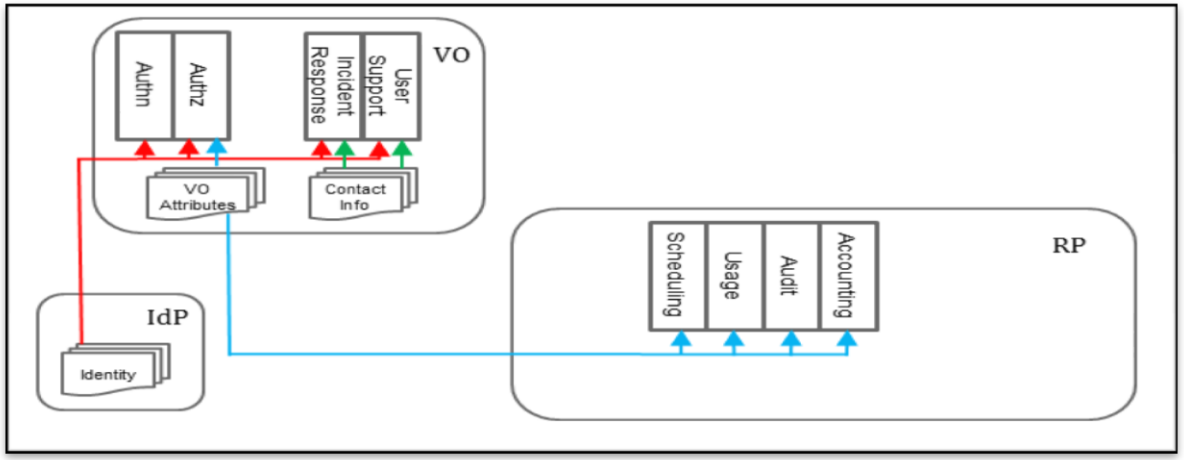
Transitive trust relationships forego all user data consumption by RP.

Virtual Organization (VO)

- Created to manage scientific community
- Role in Transitive Trust (Delegated) IdM model
  - Resource Providers (RPs) trust the VO to manage its community
  - Little or no individual user information is transferred from the VO to the RP
  - Central participant in Incident Response

Transitive Trust VO Architecture

- User registers required attributes with VO
- VO handles support and incident contacts; does not pass attributes to RP
- IdP only needs to verify authentication



Drivers & Benefits of Delegation

- Allow scaling of scientists, RPs.
- Centralize management of VO policies.
- Place effort where most appropriate.
- Avoid unneeded duplication of IdM data.
- Eases collaboration inside of and across VO.
- Improve ease of use through better integration with science workflows.

Seemingly Contradictory Demands

- Current Processes and Policies
  - Strong identification, authentication, and authorization of user communities
- User communities
  - Large scale with dynamic membership
  - Span multiple resource providers
  - Desire ease-of-use (*e.g.* single sign-on)
  - Self management

Enablers of Delegation

- RP-VO existing relationships and explicit agreements
- Available VO IT/IdM Effort and Expertise
- User traceability mechanisms (glExec)

Potential Barriers

- Historical Inertia and Risk
- Compliance and Assurance
- Technology Limitations
- Poor User Traceability Mechanisms

Historical Inertia and Risk

- Significant policy and cultural investment in current risk profile for cyber security
- DOE recognized need to shift to risk-based security with O 205.1B in 2011
- Cyber program can be flexible if risks are documented and residual risks accepted
- Transitive trust may significantly reduce costs with little increase in residual risk

Compliance and Assurance

- Stakeholders of RP tend to have higher IdM requirements
  - Strength of authentication
  - Traceability
  - Auditing, and accounting
- Some stakeholders are realizing that persistence and contact information are more important

Technology Limitations

- Many tools (source code systems, ssh, *etc.*) assume traditional authentication
- Technology advances are coming rapidly
  - Virtualization
  - Grid and cloud computing
- Increased ability to share resources within a group and increase isolation and security from other groups

Traceability

- Throughout history of LHC grid, this has been a requirement by the RPs
- With transitive trust, RP has no ability to contact individuals
- OSG Traceability Project investigated– except in improbable circumstances, sufficient information is available
- Security issues if all jobs run with same userid

Issues w/ Transitive Trust (Delegated) IdM Model

- Lack of persistent storage at RP
- VO-provided services expect user identities)
- User Support and Incident Response coordination

Roadmap for Incremental Implementation

- Delegation of IdM is not all-or-nothing
- Partial delegation – certain functions – can create a simpler workflow (for RPs and users)
- Trusting the VO and accepting the risk can significantly decrease administrative costs

Actionable Guidance for the US DOE Community

Robert Cowles, Craig Jackson and Von Welch.
Facilitating Scientific Collaborations by Delegating Identity Management: Reducing Barriers & Roadmap for Incremental Implementation

- http://cacr.iu.edu/sites/cacr.iu.edu/files/FSCbyDIM0408.pdf
- Was presented at CLHS 15, June 2015

Discusses Deemed Export, Unclassified Foreign Visits, clouds to overcome technology limitations and increase isolation, and shift to risk-based security to overcome inertia.

Lessons Learned in the Course of This Research

- Innovations
- Challenges

Innovations

- Composition of the team
- Emphasis on knowledge rather than code
- Comprehensive and comprehensible model
- Evidence-based research

Challenges

- Engaging collaborations at the right time
  - Too early – not engaged in design
  - Too late – IdM design often complete
- Reaching the right target audience for papers and presentations

More Work is Required ….

- Remaining drawbacks
- Integration of IdM into supporting infrastructures
- Developing a taxonomy of scientific data and their security requirements

Remaining Drawbacks

- Privacy constraints
  - IdP may not want to release attributes
  - Users may not attributes released widely
- Long tail of science
  - Small VOs – lack IT infrastructure
  - No IdM and little IT expertise
  - Complex user interface
- Lack of trust in small VOs by RPs

Scientific eCollaborations (SeCs)

- SeCs are initiated in a framework tailored to particular scientific disciplines
- New members register with the SeC and provide contact information
- Members authenticate with federated credentials
- There is no requirement for “release of attributes” from the IdP since the collaboration knows who it is admitting

Vision – SeCs

- Resource providers have no need to know the identity of a user; access control roles and rules are determined by the VO; eliminates need for RP to maintain an IdM for the VO
- In the case of account compromise at an IdP, RPs are protected as soon as the SeC disables the member's access rights until the IdP takes care of the issue

Operation – Critical to Success

- A trusted entity to provide the IT operations for the VOs
- Ensure appropriate Incident Response
- Provided not as an Infrastructure, not as a Platform, but as a Service (SeCsaaS)
- Organizations like EGI, GÉANT, and/or OSG might be appropriate to sponsor or actually operate SeCs

Taxonomy of Data vs Security

- Varying requirements from various sources
- Uncertainty in sensitivity of data leads to conservative security decisions
- May inhibit collaboration and discovery
- Comprehensive and comprehensible framework of requirements will realty ease task of decision makers

Thank you. Questions?

Bob Cowles (bob.cowles@gmail.com)

https://cacr.iu.edu/collab-idm

We thank the Department of Energy Next-Generation Networks for Science (NGNS) program (Grant No. DE-FG02-12ER26111) for funding this effort.