

# Identity Management in Future Scientific Collaborations (XSIM)

---

*Bob Cowles, Craig Jackson, Von Welch (PI)*

*HEPiX Fall 2013 Workshop  
University of Michigan, Ann Arbor  
October 30<sup>th</sup>, 2013*



**CENTER FOR APPLIED  
CYBERSECURITY RESEARCH**

---

INDIANA UNIVERSITY  
Pervasive Technology Institute

# HISTORY OF SCIENTIFIC COMPUTING IDENTITY MANAGEMENT



**CENTER FOR APPLIED  
CYBERSECURITY RESEARCH**

INDIANA UNIVERSITY  
Pervasive Technology Institute

# Identity Management (IdM)

From Wikipedia: “**Identity management** describes the management of individual identifiers, their authentication, authorization, and privileges within or across system and enterprise boundaries with the goal of increasing security and productivity while decreasing cost, downtime and repetitive tasks.”

Who users are and what they can do.

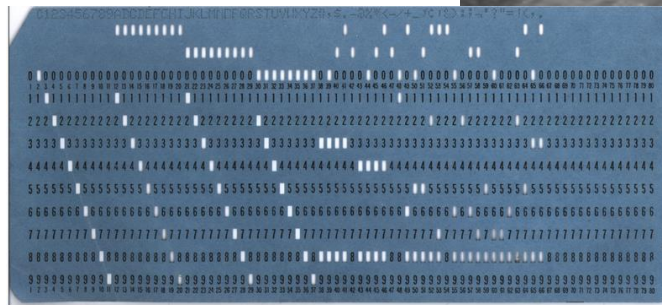


# At first, the scientist went to the computer.

Scientists were employees or students of the resource provider.



*Image credit: Lawrence Livermore National Laboratory (via Wikipedia)*



*Image credit: Wikipedia*



**CENTER FOR APPLIED  
CYBERSECURITY RESEARCH**

INDIANA UNIVERSITY  
Pervasive Technology Institute

# Growth of the collaborations

Number of scientists, institutions, resources.  
Large, expensive, rare/unique instruments.  
Increasing amounts of data.

The model of a single resource provider managing all their users started eroding.

## Some history of scale...

Date	Collaboration sizes	Data volume, archive technology
Late 1950's	2-3	Kilobits, notebooks
1960's	10-15	kB, <u>punchcards</u>
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*

# Enter the collaboratory



The collaboratory has proven itself as the key way of allowing multi-organization science collaborations to utilize a wide variety of resource providers. We now have 15 years of applied experimentation in how collaboratories implement IdM.

ATLAS: 3,000+ members, 177 institutions, 38 countries.

CMS: 3000+ members, 172 institutions, 40 countries.

ALICE: 1200+ members, 132 institutions, 36 countries.

XSEDE: 10000+ users, 16 resources.

LIGO: 800+ scientists, 56 institutions, 13 countries.

Etc.



# XSIM Vision

Enable the next generation of trustworthy extreme-scale scientific collaborations by understanding and formalizing a model of identity management that includes the collaboratory.



# XSIM Approach

Determine the motivations that lead to different choices and develop a Collaboratory-IdM **model** to express the **trust** relationships between resource providers (RPs) and current (based on **interviews**) and **future** collaboratories.

Validate the model and develop guidance to collaboratories and resource providers in architecting their IdM and trust choices.

# First Step: Define Trust Relationship

Large body of research on trust exists, in computer security, CS, and more broadly, but no clear consensus on definition.

We looked at many and settled on:

## *Trust* –

A disposition willingly **to accept the risk of reliance** on a person, entity, or system to act in ways that benefit, protect, or respect one's interests in a given domain.

Based on Nickel & Vaesen, Sabine Roeser, Rafaela Hillerbrand, Martin Peterson & Per Sandin (eds.), *Handbook of Risk Theory*. Springer (2012)



# Interviews

GOAL: Understand the trust relationships (accepted risks) among resource providers/collaboratory/users and how those were arrived at.

Key to understanding the “real reasons” behind implementation and lessons learned.



# Model Basis: Collaboratory IdM Lifecycle

- Enrollment
- Provisioning
- Request
- Usage
- Incident Management
- De-provisioning

A common IdM concept.

Each lifecycle stage has a small number of possible collaboratory/RP interactions.

First exposure of user-specific information is a big one.



# Simple Version of Model

- Primarily:  
When (lifecycle stage) does user-specific information flow from collaboratory to RP?
- “When” is expressed in collaboratory lifecycle:  
Enrollment, provisioning, request, usage, user support/incident response, (de-provisioning,) never.



# What Does The Model Mean?

- Early identification of user by collaboratory to RP => less delegation and trust by RP. At extreme, the collaboratory is just an interface to RP.
- Later/no identification of user by collaboratory to RP => either:
  - More trust of collaboratory by RP; or
  - Desire of RP to have less effort.

# More Refined:

## Factors Affecting IdM Design

- User-user Isolation
- Persistence of user data or state
- Complexity of collaboratory roles
- Scaling in terms of collaboratory users
- Incentive balance: collaboratory <-> RP
- Inertia – early relations more conservative
- Technology limitations

# FUTURE RESOURCE PROVIDERS



**CENTER FOR APPLIED  
CYBERSECURITY RESEARCH**

INDIANA UNIVERSITY  
Pervasive Technology Institute

# HPC

- Service parallel tasks with requirements for low latency communication (often using InfiniBand) and high-end processors
- Relatively small number of users conforming to site-specific infrastructure
- Shell access increases security considerations (vetting, 2-factor, etc.)

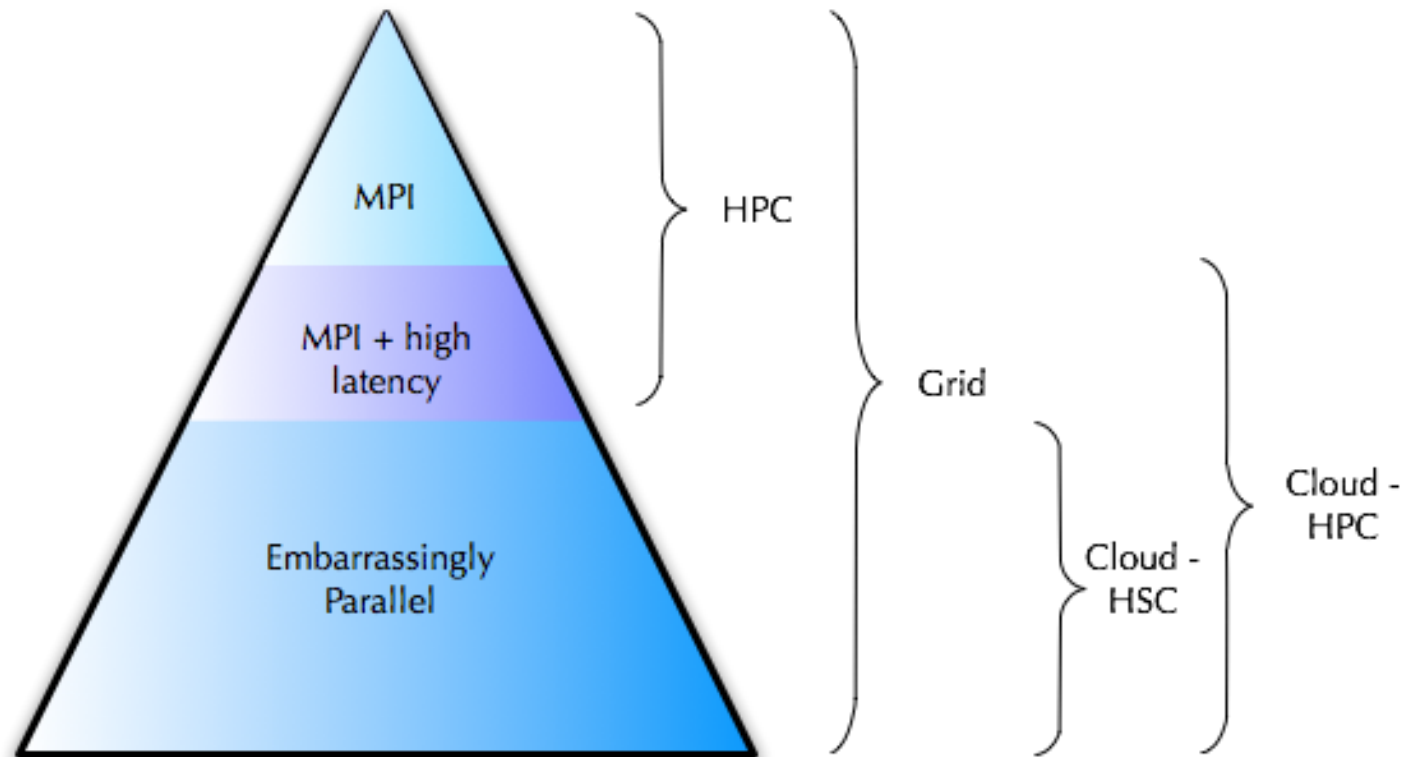


# Cloud

- Highly Scalable Computing (HSC)  
Embarrassingly parallel  
Commodity processors  
Relatively little process coupling
- On-demand access to homogenous virtual resources
- Private (enterprise) and public (commercial) implementations



# HPC vs. HSC (plus Grid)



<http://www.cloudscaling.com/blog/cloud-computing/grid-cloud-hpc-whats-the-diff/>

# Cloud – Survey results

- Benefits

  - Don't have to fit into existing infrastructure

  - Elasticity in compute and data

- Challenges

  - Requires IT expertise

  - Lack of cloud interoperability

  - Data (security, stability, bandwidth, file systems)

  - Funding

Source: XSEDE Cloud Use Survey presented at EGI-TF 2013

# Federations - HPC

- XSEDE in US and PRACE in Europe provide for increasingly seamless use of HPC clusters
- Support for limited number of common frameworks allows for some flexibility and interoperability
- Portals help hide UI complexity



# Federations - Cloud

- Private – Public federations  
Integrate with a specific commercial cloud provider to enable response to peak demands
- Private – Private federations (research)  
Integration across domains providing researchers access to broad range of resources

# FUTURE COLLABORATORIES



**CENTER FOR APPLIED  
CYBERSECURITY RESEARCH**

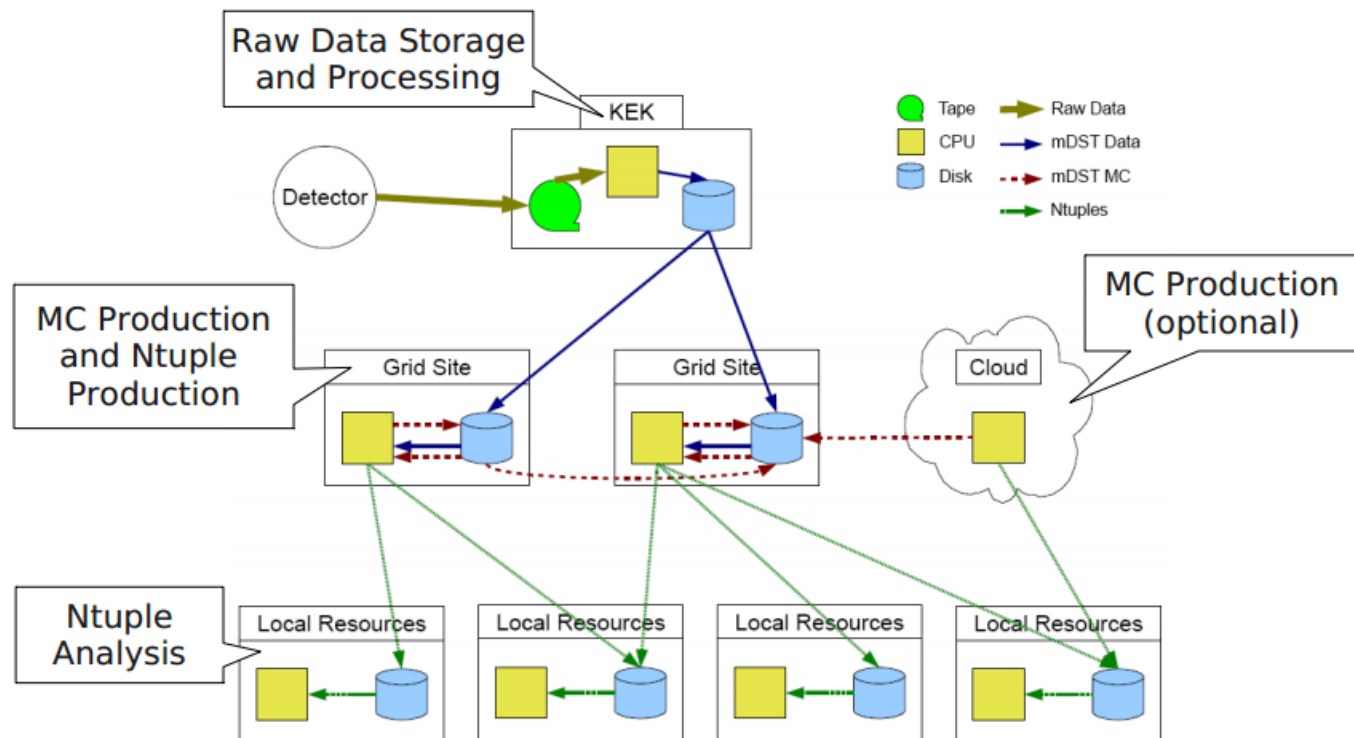
INDIANA UNIVERSITY  
Pervasive Technology Institute

# HEP

- LHC (ongoing – computing model updates planned to accommodate the significant luminosity increases in 2022 timeframe)
  - ATLAS
  - CMS
  - Alice
  - LHCb
- Belle-II (start-up in 2015 – newest large HEP collaboration)
- ????



# Belle-II Computing Model



*Courtesy of Thomas Kuhr - KIT*

# Sample Areas of Future Growth

- Astrophysics
  - Sky surveys (FST, DES, SKA, LSST)
  - Dark energy, dark matter
- Biomedical – Genomics, Pharmaceuticals
- Chemistry – reactions, materials science
- Earth Sciences – Climate modeling
- Physics – Gravity, WIMPs

# Significant Differences from HEP

- Large data sets with non-independent events
- Security and privacy data issues
- Distributed data sources
- Distributed IT support infrastructure  
eLog, Wiki, analysis portals, admin
- Lack of IT expertise

*“Long tail of Science”*

## Challenge(s)

What forms of identity / attribute management can better serve the requirements of the broad scientific community?

Are there legal issues to address?

Are there policy issues to address?

Are there security issues to address?

# XSIM FUTURE WORK



**CENTER FOR APPLIED  
CYBERSECURITY RESEARCH**

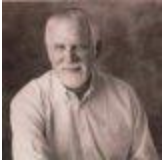
INDIANA UNIVERSITY  
Pervasive Technology Institute

# Future Work

- More diverse resource providers and collaboratories: exascale, cloud, “long-tail of science”
- Implications of trust violation.
- Better understand motivations to create guidance for new collaborations.
- Apply model with real-world collaboratories and within the Open Science Grid.



# The XSIM Team



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



**Craig Jackson** – CACR Policy Analyst, former practicing attorney.



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

The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the sponsors or any organization.

# Thank you. Questions?

Bob Cowles

([bob.cowles@brightlite-infosec.com](mailto:bob.cowles@brightlite-infosec.com))

<http://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.

