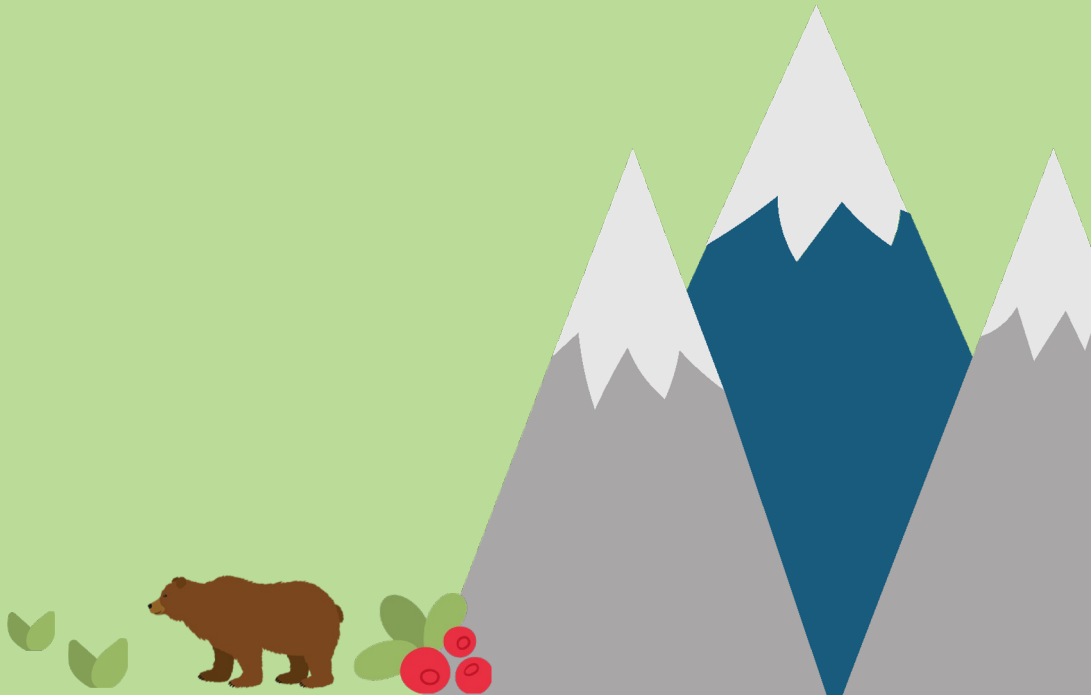
Data in Practice: Understanding the Role of Data in IAM

Chris Bongaarts, IAM Architect | University of Minnesota
Kellen Murphy, Identity Architect | University of Virginia

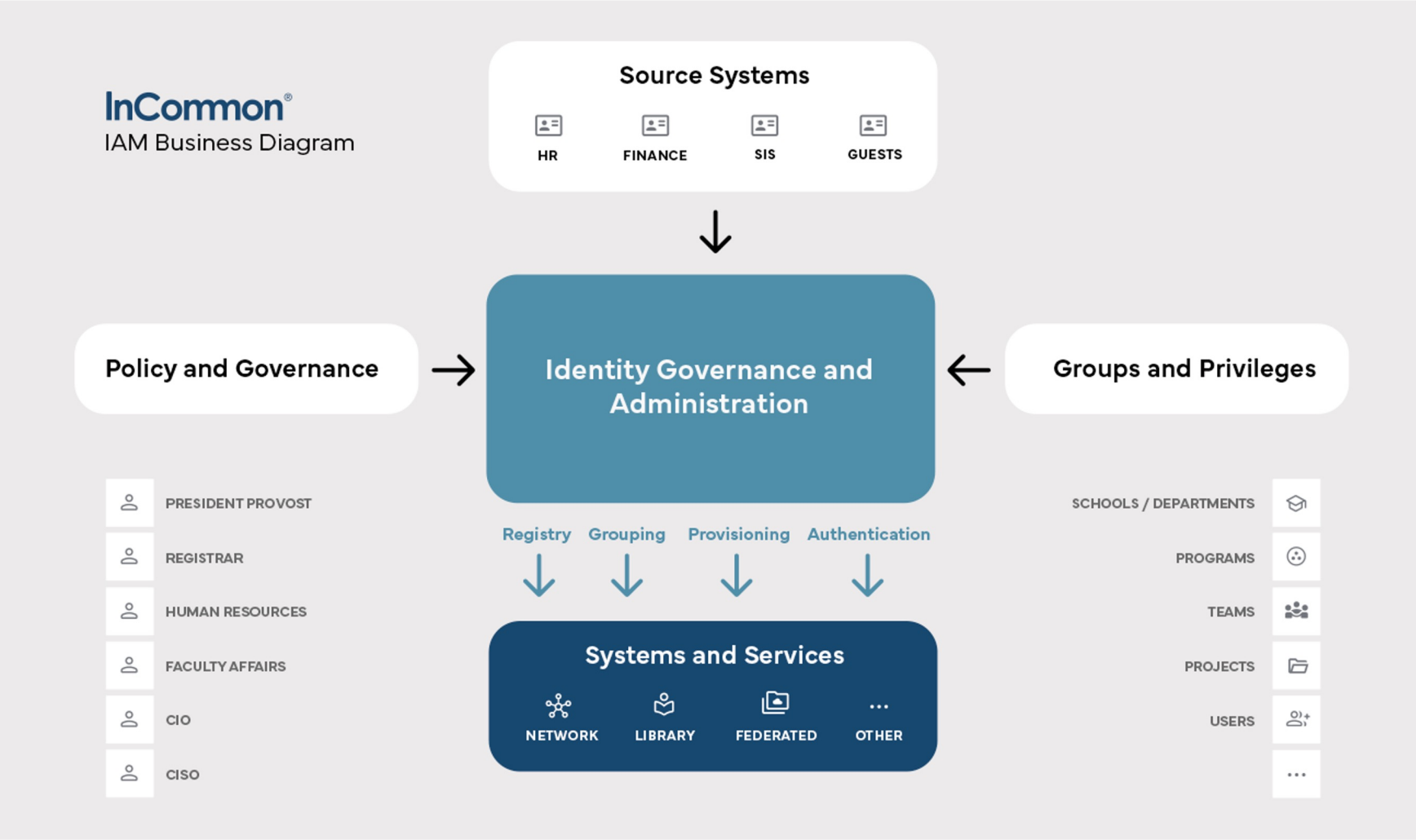
Date: June 5, 2025



Agenda

- Common Data Sources
- How IAM Systems use Data
- Problems from Incomplete Data
- Problems from Overcomplete Data
- Data Flow
- Data-based Decision Making

Context reminder: The big picture



Also this one...

Common Data Sources

- Systems of Record
  - Student Information System
    - Registration (courses, terms, majors), Admissions
  - Human Resources
  - Alumni Association/Foundation
  - Other colleges/departments
  - IAM system
    - "guest" accounts, functional accounts, “sponsored” accounts, etc.
  - Social Identity Providers
- “Identity First”
  - IAM is the SOR for Identity Data

How IAM Systems use Data

- Identity Lifecycle
  - Attributes – Data about YOU.
    - … from a System of Record
      - cn, sn, givenName
      - University ID
    - … from the Identity System.
      - pwdLastReset
  - Roles – Group of individuals with a similar set of attributes.
    - faculty
    - staff
    - student
    - etc.
  - Entitlements – Privileges you get…
    - Because of your role…
    - Because you asked for them / need them…

Examples: Data @ UMN

Attribute Description	Value
objectClass	***umnPerson (structural)***
cn	**Christopher A Bongaarts (cab)**
sn	**Bongaarts**
displayName	Christopher A Bongaarts
eduPersonOrcid	0000-0002-4225-9796
eduPersonPrincipalName	cab@umn.edu
facsimileTelephoneNumber	+1 612-626-[redacted]
freeFormName	Christopher A Bongaarts
givenName	Christopher
homePhone	+1 952-[redacted]
homePostalAddress	[redacted] Burnsville, MN 55337-[redacted]
info	Summer 2003
initials	A
l	Minneapolis
labeledURI	http://umn.edu/lookup/cab Christopher A Bongaarts
mail	cab@umn.edu
o	UMN Twin Cities
ou	Identity Access Management
pager	+1 612-[redacted]
physicalDeliveryOfficeName	Room 62[redacted]
postalAddress	Room 62[redacted]
postalCode	55454
preferredRfc822Recipient	cab@[redacted]
serialNumber	R162[redacted]
st	MN

Sizing Information	
Entry Size:	8 KB (8448 Bytes)
Number of Children:	Not checked
Number of Attributes:	73
Number of Values:	210

<- Attributes

Person types ("roles") ->

Person type entitlement matrix -v

FPERSON TYPE	DESCRIPTION	IDM TYPE	ACCOUNT SOURCE
POI-regent	University Board of Regent member (Non-Employee). Central HR adds the POI-regent type.	Dynamic	Peoplesoft HR Solutions: PS_HR
Proxy	Parent/Guest Access - These are people who have access to Myu, so that they can view and pay student bills, view grades, view class schedule, and view housing	Dynamic	Peoplesoft Campus Solutions: PS_CS
recentStudent	Enrolled for the term prior to the current - Exclude summer semester in this logic	Dynamic	Peoplesoft Campus Solutions: PS_CS
retStaff	Employees who have retired from U of M.	Dynamic	Peoplesoft HR Solutions: PS_HR
Friend	Sponsored - An employee at UOM sponsors this person.	Static	OIM
Staff	Current, Real and a Non Student Job. Also all active POI's are staff	Dynamic	Peoplesoft HR Solutions: PS_HR
Student	Enrolled as a student at some time	Static	Peoplesoft Campus Solutions: PS_CS
StudentOrg	Student can create their own		

Account type	DUO Required	VPN	Wireless	UCard	Active Directory	Google Apps	Google Apps EDU Plus	LDAP
HR Accounts	Account information source is PeopleSoft HR and is controlled by Human Resources							
Staff	Y	Y	Y	Y	Y	Y	Y	Y
StudentStaff	Y	Y	Y	Y	Y	Y	Y	Y
futureStaff	Y	Y	Y	Y	Y	Y	N	Y
nonStaff	Y	N	N	N	N	N	N	N

Managed Outside of the Identity System

Account type	Zoom	MyU	Office365 / Adobe Acrobat Pro DC	TDX
HR Accounts	Account information source is PeopleSoft HR and is controlled by Hum			
Staff	Y	Y	Y	Y
StudentStaff	Y	Y	Y	Y
futureStaff	N	N	N	N
nonStaff	N	Y*	N	N

Examples: Data @ UVA

- At

Update or view user profile details.

Legal First Name:	Thomas	Legal Middle Name:		Legal Last Name:	Jefferson
Computing ID:	tj4u	University ID:	743008857	Name Suffix:	
Preferred First Name:	Thomas	Preferred Middle Name:		Preferred Last Name:	Jefferson
Display Name:	Thomas Jefferson	UVA Display Name:	Jefferson, Thomas	Preferred Generational Suffix:	
National ID (SSN):		Birth Date:	4/13/1826	More Important:	
Address:		Home Street 2:		Home Street 3:	

- Roles (20 values)
  - student / faculty / staff
    - applicant / pre-
    - former / alumni
    - UVA vs College at Wise
    - Health System
- Entitlements (1445… and counting)
  - Set various parameters in various systems, e.g. when does a particular attribute get set to ‘true’

Roles - Private LDAP

Role	Exist in Private LDAP	UVAActiveStatus	UVA_Member	UVAPERSONIAMAFFILIATION	eduPerson Schema
Student Applicant	Y	InActive	No	student_applicant	–
Student	Y	Active	Yes	student	Student
Student (Grace)	Y	Active	Yes	grace_student	Student
Former Student	Y	Active	Yes	former_student	Student
Alumni	Y	Active	Yes	alumni	Alum
Faculty	Y	Active	Yes	faculty	Faculty
Faculty (Grace)	Y	Active	Yes	grace_faculty	Faculty
Faculty Emeritus	Y	Active	Yes	emeritus	Faculty
Staff	Y	Active	Yes	staff	Staff
Staff (Grace)	Y	Active	Yes	grace_staff	Staff
Former Employee	Y	Active	Yes	former_employee	–
Retiree	Y	Active	Yes	retiree	–
Sponsored Account	Y	Active	Yes	sponsored	Affiliate
Wise Student	Y	Active	Yes	wisestudent	Student
Wise Student (Grace)	Y	Active	Yes	grace_wisestudent	Student
Pre-Staff	N	–	–	–	–
Pre-Faculty	N	–	–	–	–
Student Worker	Y	Active	Yes	student_worker	Staff
Student Worker (Grace)	Y	Active	Yes	grace_student_worker	Staff
Community Health	Y	Inactive * if only Community Health	Yes	communityhealth	–

UVA_MEMBER = NO if no role
UVASTATUS = Inactive if no role

Problems from Incomplete Data

- An IAM system is only as good as the data it receives.

Good Data

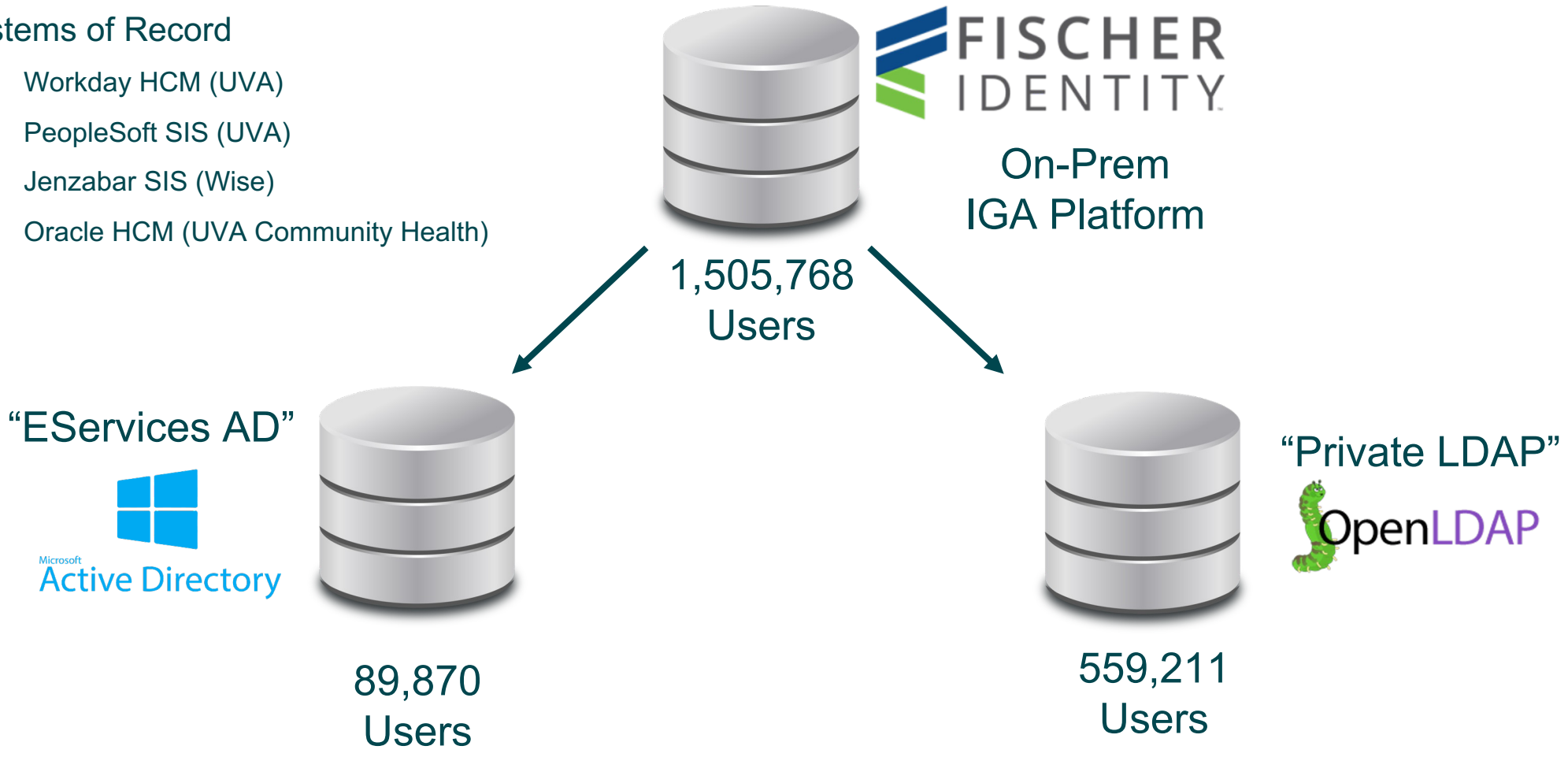
- The minimally required accounts are issued.
- Those accounts are only granted the privileges they need.
- Compliance: reports are accurate and complete.

Bad Data

- Users are issued duplicate, inaccurate, or missing accounts as a result of mismatched or incomplete data.
- Users given excessive or fewer privileges than they should have.
- Disparities in identity records contributing to regulatory infractions.

Examples: *Incomplete* Data @ UVA

- Systems of Record
  - Workday HCM (UVA)
  - PeopleSoft SIS (UVA)
  - Jenzabar SIS (Wise)
  - Oracle HCM (UVA Community Health)



Examples: *Incomplete* Data @ UVA

- Systems of Record
  - Workday HCM (UVA)
  - PeopleSoft SIS (UVA)
  - Jenzabar SIS (Wise)
  - Oracle HCM (UVA Community Health)

FISCHER IDENTITY

On-Prem IGA Platform

1,505,768 Users

PROBLEM

Data mismatches regularly occur.

We *don't* update users outside of SORs, even though SOR data is wrong / incomplete by policy.

Upstream issues are regularly "blamed" on IGA despite not originating in Fischer IGA.

"EServices AD"

Microsoft Active Directory

89,870 Users

PROBLEM

Not all of the users that really need to be in the directory are there… e.g. big issues for Entra ID SSO

559,211 Users

"Private LDAP"

OpenLDAP

PROBLEM

Many users in a state inconsistent with their current role in IGA… need to "replay" users in batches regularly.

Problems from Overcomplete Data

- Consider your "Total Cost of (Data) Ownership"
  - Acquiring the data (and maintenance of that process)
  - Storing the data safely
  - Risks of inadvertent disclosure (breach notification)
  - Data you have may be legally discoverable
- Mitigation Strategy: Minimization
  - Only ask for/store what you NEED
  - Eliminate what you don't (e.g. SSN abatement)
  - "Shift left" where you can (SOR handling identity match/merge/split)
- Just-in-time vs. Just-in-case
  - Tradeoff: Data maintenance vs. resiliency
  - Use "caching" sparingly where engineeringly appropriate

Examples: *Overcomplete* Data @ UMN

- Long history of IAM (1992)
  - Acquired data in a very different environment (mainframe, separate HR vs SIS)
  - Today: Peoplesoft (shared person data), higher security awareness
  - IAM was data broker of choice for a long time
    - Now trying to get out of that business
- Old IAM system retirement
  - Rebuilt feeds with new data sources
  - Not all data available there
  - Derived data not necessarily easy to replicate (transforms in and out; population)
  - Example: library privileges depending on data (campus) for retired staff no longer in HR

Data Flow

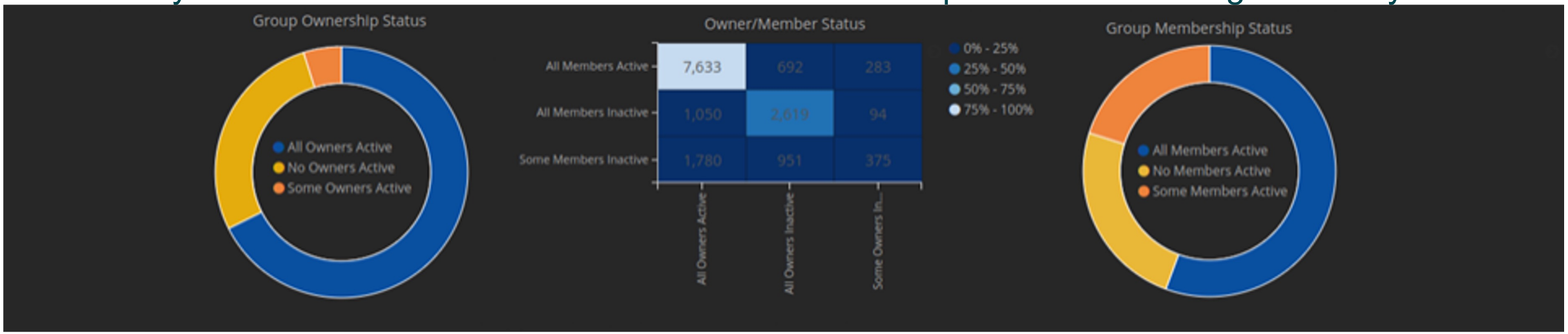
- Upstream vs downstream flow
  - IAM is usually in the "middle"
  - Some data may flow back to systems of record (e.g. centrally assigned email address)
    - But beware of data "cycles"
  - IAM often provisions major applications (email, directories)
  - IAM may be a place to add data that source systems don't support
    - Preferred name, pronouns
    - ORCID iD
  - Track who you give data to, what you give them, and WHY
    - Ideally this would be part of a more general Data Governance program

Data Flow

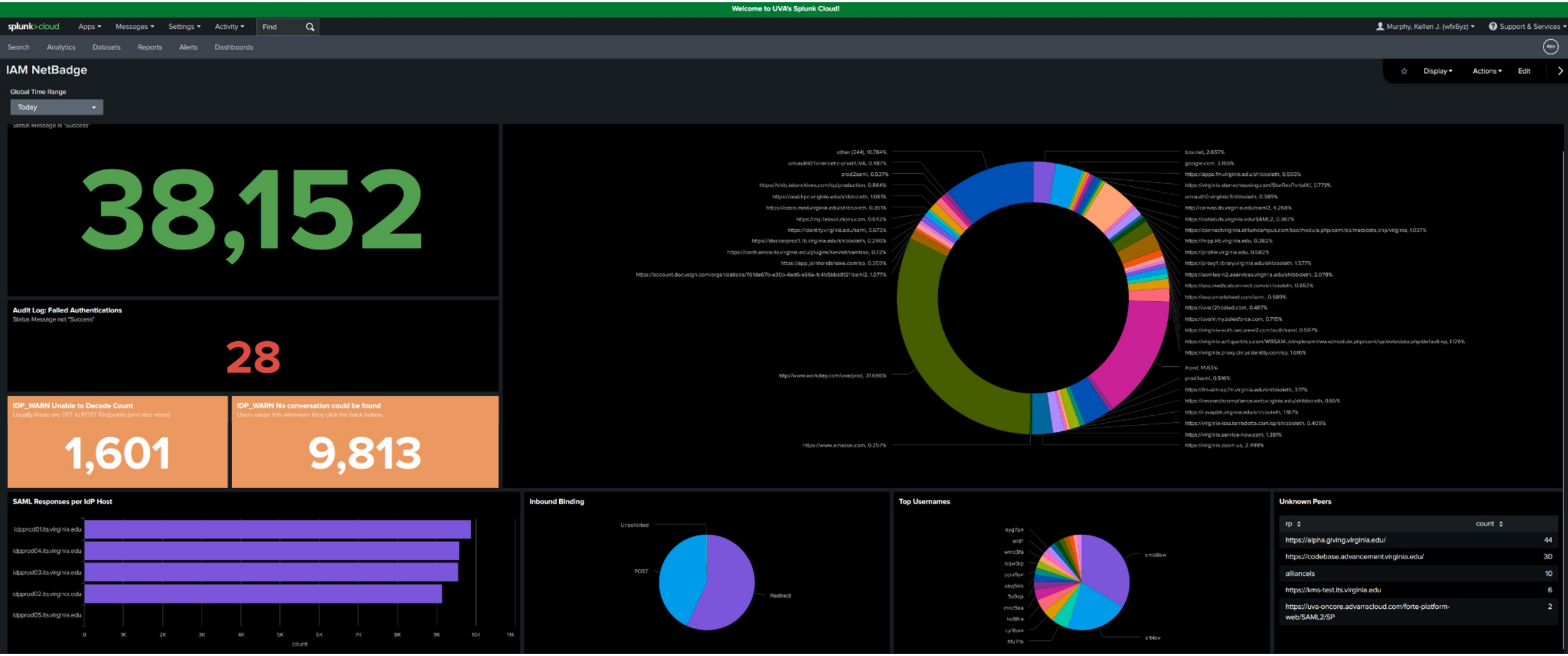
- Incorrect data should be fixed at the source
  - Saves effort of correcting in many places
  - Saves effort of recorrecting in many places when source data is reapplied
  - It is worth building strong **relationships with upstream data owners/custodians** to facilitate this
- Engineer for peak data change (know your business cycles!)
  - Beginning, end of term ("future student" -> "current student" -> "recent student")
  - HR annual contract renewals (lots of notifications but not to data IAM needs)

Data-based Decision Making

- Identity data is natural for segmentation analysis: *who* is using *what* and *how*.
- Centralized identity data stores can unify information from across a *de*centralized organization – data is more holistic, leaders will make better decisions.
- Identity-driven data is ideal for dashboards to allow for rapid decision making and analysis:



Log Data


Welcome to UVA's Splunk Cloud!
splunk>cloud
Apps
Messages
Settings
Activity
Find
Murphy, Kellen J. (wfx6yz)
Support & Services
Search
Analytics
Datasets
Reports
Alerts
Dashboards
IAM NetBadge
Display
Actions
Edit
Global Time Range
Today
38,152
Audit Log: Failed Authentications
Status Message not "Success"
28
IDP_WARN Unable to Decode Count
Usually these are GET to POST Endpoints (and vice versa)
1,601
IDP_WARN No conversation could be found
Users cause this whenever they click the back button.
9,813
other (244), 10.784%
box.net, 2.657%
google.com, 3.165%
https://apps.fm.virginia.edu/shibboleth, 0.503%
https://canvas.its.virginia.edu/saml2, 4.268%
https://collab.its.virginia.edu/SAML2, 0.367%
https://ihrpp.irb.virginia.edu, 0.382%
https://profile.virginia.edu, 0.582%
https://proxy1.library.virginia.edu/shibboleth, 1.577%
https://somleam2.eservices.virginia.edu/shibboleth, 2.078%
https://uvahr.my.salesforce.com, 0.715%
https://virginia.proxy.cirrusidentity.com/sp, 1.616%
http://www.workday.com/uva/prod, 31.686%
https://www.amazon.com, 0.257%
SAML Responses per IdP Host
idpprod01.its.virginia.edu
idpprod04.its.virginia.edu
idpprod03.its.virginia.edu
idpprod02.its.virginia.edu
idpprod05.its.virginia.edu
0
1K
2K
3K
4K
5K
6K
7K
8K
9K
10K
11K
count
Inbound Binding
Unsolicited
POST
Redirect
Top Usernames
Unknown Peers
rp
count
https://alpha.giving.virginia.edu/ 44
https://codebase.advancement.virginia.edu/ 30
alliancels 10
https://kms-test.its.virginia.edu 6
https://uva-oncore.advarracloud.com/forte-platform-web/SAML2/SP 2

InCommon BaseCAMP

LEARN. CONNECT. ENGAGE.

Thank you!