CODEBOOK

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
USE CASES OF RC	—	*Statements highlighting the different ways researchers utilize the Research Computing (RC) infrastructure.*	"for AI related work HPC is the way to go" large-scale data processing and visualizations support secure research with HIPAA compliance run and process simulations run and scale large experiments data storage paid virual machine with full administrative access share similar works/scripts with others handle biomedical research data for documentation and project management purposes shared custom application software free resources open to all university researchers for analysis purposes "Science Gateways" share results shared software
ROLE OF DIFFERENT ACTORS	ROLE OF LOCAL ADMIN	*Distinct responsibilities of a Local Admin within the RC ecosystem*	discussions with PIs helping students with files and tools help researchers run experiments on the cluster determine resource availability as per PI request track access permissions given to researchers local admin isn't involved in the access rule decisions PI as the admin is not ideal
	ROLE OF PI	*Distinct responsibilities of a Principal Investigator (PI) within the RC ecosystem*	PI gets personal influences but no technical benefit sponsor student accounts in RC PI manages/keeps track of access access permissions decided/approved by PI collaborator granted access by PI gatekeepers of the projects
			build workforce by training students how to administer

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
	ROLE OF SYSADMIN	*Distinct responsibilities of a system administrator within the RC ecosystem*	build self-service facilities for users to manage assets build bespoke systems to support research needs maintain security of the infrastructure through os patching support researchers from diverse disciplines admins get often inquired about resource availability different resources managed by university sys admin PIs are dependent on Sysadmins to provide access researchers are largely dependent on admins work with PIs to manage privileges of sponsored accounts admins support researchers with resource sharing finding a middle ground to help researchers share data
	ACCESS MANAGEMENT	*Policies followed by Admins to govern Access and Privilege Management within RC*	"PIs are the first level controller" PI gets special access privileges in condo model maintain user roles to identify access permissions access control needs both admins and users delegating access decisions for group members to PIs group members can work out the permission model for project dir "most policies are managed at group level" PI approves access and can access all students' data user-level access stricter policies for HIPAA compliance privileges are managed differently: data vs computing resources "owners can share their private nodes" access permissions for RC as a whole from leniant to collaborative policies for resource sharing no standard system for privilege audit/tracking in place manual revocation policy non-standard revocation policy - availability vs security per-group vs per-user usage policies all regular users have same privileges privilege sharing (DAC) are restricted in home and scratch example of group-based access control for collaboration

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
ADMINISTRATIVE POLICIES			sysadmins dont enforce fine-grained rules on private resources PI can delegate responsiblity to other users admins cannot do low-level enforcement for all user activities admins only monitor user activities for policy compliance
	USER AND GROUP MANAGEMENT	*Policies followed by Admins to govern User and Group Management within RC*	PI controls how they want to manage their group different types of user accounts "users can create their own groups" groups are managed using PI's credentials students require a PI sponsor to access RC PIs remain a root of trust as they sponsor user accounts admins are not invloved beyond group-level
	RESOURCE MANAGEMENT	*Policies followed by Admins to govern Resource Management within RC*	owned resource allocation vs type matching hotel model of sharing resources condo model of sharing resources all that RC manages is "public," condo too when idle computing resources segmented into access zones homogeneous vs heterogeneous resources mngng resource alloc policy is tricky - fairness vs utilization example of administrative policies on resource allocation sysadmins have the right to delete scratch data
	OTHER POLICIES	*Other miscellaneous comments around RC Management Policies*	PIs have more say/influence on their owned resources having formal administrative policies can help in long run funding decides the nature of the cluster goverance board with broad perspective difference between industrial and academic rc
	ACCESS MANAGEMENT	*Technical processed and procedures around Access and Privilege Management within RC*	automated alert for unauthorized privilege escalation PIs manage sponsor accounts permissions via a web-based tool file permissions managed using unix permissions and groups Fair share model for resource allocation FCFS for resource allocation in private partition user can change directory persmissions Globus is implemented for global sharing

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
TECHNICAL PROCESSES AND PROCEDURES			data-based access permissions scratch space open to ALL uses PAM to restrict job allocation
	USER AND GROUP MANAGEMENT	*Technical processed and procedures around User and Group Management within RC*	rc account creation process user attributes are used for usage tracking and accounting comprehensive user attributes sys admin perspective of managing user and groups in RC system supports multi-sponsor accounts and cross germination QoS helps differentiate between users
	RESOURCE MANAGEMENT	*Technical processed and procedures around Resource Management within RC*	data ownership is imp; more security controls for on prem data ad-hoc process for maintaining resources resource attributes resource orchestration process
ADMINISTRATIVE PRIVILEGES	UNDERSTANDING OF ADMINISTRATIVE PRIVILEGES	*Participants' comprehension of administrative privileges within the context of RC*	root privileges demand a lot of responsibility PI training is crucial before granting admin privileges "if you manage to be root, then you're God" "users are not allowed to run as root anywhere" more system users at root level, higher the rate of error admin access is never provided for remote file systems admin privileges are requested due to unawareness of processes permission mgnt for partitions is difficult for PI to handle admin oversight is imp for groups to prevent any misuse requirements of admin access can be resolved via consultation Individual sharing without admin oversight is hard to manage Giving admin prvlgs to PIs can lead to errors
	CANDIDATES OF ADMINISTRATIVE PRIVILEGES	*Opinions and wishes on who should have administrative privileges and why, considering specific contexts and purposes*	RC should delegate file system permissions to users users (researchers) of private resources are good candidates administrative privileges at PI level is wished admin privileges to a representative from lab is good prvlgs affctng whole infrastructure should be kept to sysadmins
			forming new generations of researchers through mentorship external collaborations with experts/other institutions

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
COLLABORATIVE RESEARCH TEAMS	REASONS FOR COLLABORATION	*Reasons why researchers choose to collaborate within academic settings*	academic collaborations are eternal collaboration with people from similar backgrounds exchange of skills/resources delegation of responsibility working toward the same goal create harmonious datasets
	COLLABORATIVE TEAM STRUCTURES	*Different ways collaborative research teams are organized*	members with distinct responsibilities/stake interdisciplinary collaborative team single handedly led projects research teams built on trust members with different academic levels matrix-kind of interrelationship team structures primary roles in academic research teams - lead student & PI hierarchical vs disordered team structures contributions-based negotiations for project leads lead author calls the shots
	MECHANISMS OF RESOURCE SHARING (NON-COMPUTING RESOURCES)	*Statements concerning how various non-computing resources are shared among collaborators*	trust b/w collaborators decides the nature of data sharing outsourcing public research data to journals to maintain data sharing use of commercial tools for data sharing ethical considerations and trainings challenges with using compliant vs accustomed tools tools used for resource sharing / collaboration data sharing process with external collaborators security should be considered when sharing data externally challenges with sharing large datasets using commerical tools everyone can access data sets; trusting to not misuse sharing and syncing data setsconsume time "unsolved problem"
			sharing process with multiple PIs involved sharing compute resources with ext collabs isnt straightforward sharing the idle cycles of private partition with public preferential access for certain nodes on cluster

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
RESOURCE SHARING	MECHANISMS OF RESOURCE SHARING (COMPUTING RESOURCES)	*Statements concerning how various computing resources are shared among collaborators.*	dataset size demads private RC requirements project based private resource requirements urgency/priority determining private resource sharing self-management of in-house RC resources team decides access for external member team wide concensus for resource sharing ext collabs have their own computing resources RC access to external collaborators (courtesy affiliate) PI acts as proxy for external collaborators team wide access to sponsored resources/private partition challenges with sharing RC resources with interns high prioritiy for compute hours on sponsored nodes (private) responsibility-based sharing of computational resources need to request compute time on funding agency's RC
	TYPES OF RESOURCE SHARING (NON-COMPUTING RESOURCES)	*Various types of non-computing resources shared among collaborators within scientific collaborations*	data sharing protocols prohibit data sharing w ext collabs custom software for research groups non-RC owned resources shared RC data storage
	TYPES OF RESOURCE SHARING (COMPUTING RESOURCES)	*Various types of computing resources shared among collaborators within scientific collaborations*	expertise decides sharing or not sharing project-specific compute hours special access grant for Cloud resources unaware if university RC resources are shared internal computational resources not shared with ext collabs private resources shared very infrequently externally non-RC owned resources in-house computing resource shared with collab private partition public partition exclusive access partition national supercomputers used when collaborating sharing private nodes with ext. is not possible not sharing computing resources with ext collaborators

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
FACTORS GOVERNING ACCESS PERMISSIONS	POLICY-DRIVEN FACTORS	*Policies, guidelines, or regulations that dictate access control decisions*	university or funding agencies policies common policy for public datasets
	TEMPORAL FACTORS	*Time-based factors that affect the access control decisions*	duration of the computing job permissions determined when stdnt joins/leaves lab/institut access permissions determined at the beginning of the project
	RESOURCE SPECIFIC FACTORS	*Unique characteristics of protected resources that dictate access control decisions*	sensitivity of data only (subsets of) datasets shared with collaborators data sharing should be more constrained than computing hardware-based decision making process no (raw) data shared with collaborators no large-scale data; csually sharing data w/ collaborators dataset is made open as soon as paper is published
	USER SPECIFIC FACTORS	*Unique characteristics of users that dictate access control decisions*	nationality of the researcher intellectual contributions experience of the researcher longetivity of the researcher group size and project count project and roles core team [RC] access permissions (in)different trtmnt to intrnl/extrnl users of RC arbitrary per-need basis project, roles, and responsibility based
	NOT KEEPING TRACK (WHY?)	*Statements explaining why there does not exist any implemented processes around tracking access privileges and sharedd resources within research teams*	"it's too complicated. I'm not admin" no system to track shared resources in university rc resources are limited per-user mutual trust system among team members no auditing system to track the sharing of self-managed servers enough resources that tracking is avoided trust system based access management tracking PI does not remember
			online web-based self-service portal to track allocations files/data tracked through SQLite database

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
TRACKING SHARED RESOURCES AND PERMISSIONS	KEEPING TRACK (HOW?)	*Implemented processes around tracking access privileges and shared resources within research teams*	group size and project count determine access tracking methods version control maintained by using Google Docs shared resources tracked using spreadsheet queue system to manage jobs manual process
	KEEPING TRACK (WHAT?)	*Statements describing what all information is tracked within research teams*	use cases of tracking resources amount of user information stored for resource tracking individual resource information tracked
	TRACKING LIMITATIONS	*Statements explaining why there does not exist any implemented processes around tracking access privileges and shared resources within research teams*	challenges in managing an up-to-date list tracking list is non-functional or out of date takes a lot of time to get access
	OTHER COMMENTS	*Other comments regarding tracking access privileges and shared resources within research teams*	understanding the imp of tracking resources tracking of rules and permissions is important automated access tracking will increase S&P unaware of "good" ways to track
LIFECYCLE OF ACCESS PERMISSIONS	ACCESS PERMISSIONS NOT REVISITED (WHY?)	*Statements explaining why access privileges are not revisited or updated within research teams*	no revocation policy for internal clusters
	CHALLENGES WITH NO REVOCATION	*Statements mentioning the challenges for not having revocation of access*	in instituitional RC in private RC/cluster
	ACCESS PERMISSIONS REVISITED (HOW?)	*Statements describing how access privileges are revisited or updated within research teams*	for access revocation, PI contacts RC team time based access revocation opening and closing access for data sharing purposes violation of university policy determines revocation PI authorization is needed to revoke RC access
	ACCESS PERMISSIONS REVISITED (HOW OFTEN?)	*Statements describing how often access privileges are updated within research teams*	access permissions not visited frequently regular/frequent visits for permissions national supercomputers revisit yearly
	ACCESS PERMISSIONS REVISITED (WHY?)	*Statements explaining why access privileges are revisited or updated within research teams*	annual audit for allocation renewal access records revised to grant or revoke permissions unethical access can result in permission revocation Imp for PI to act on revocation to avoid misuse
			students under the same PI can access each other's data

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
ACCESS CONTROL RULES	ACCESS CONTROL RULES IN PLACE (WHAT?)	*Statements mentioning that there exists implemented rules and policies around governing and enforcing access privileges within research teams*	two roles - Students have "user" role, PI "manager" role ext collabs can only access files, not the cluster collaborators have access on a need-to-know basis access control rule example team wide concensus for access level decisions Two access levels general rule - students have access to project resource & data whoever and whatever PI authorizes
	ACCESS CONTROL RULES NOT IN PLACE (WHY?)	*Statements explaining why there are no implemented rules and policies around governing and enforcing access privileges within research teams*	raw data is not shared with ext collaborators simulations do not need access control rules nothing sensitive relying on linux system a need for systematic access regulation hasn't arose currently no set infrastructure set up has worked so far "minimizing bureaucratic friction" lacking personnel with management experience using built-in access control mechanisms for cluster reliance on trust
	OTHER COMMENTS	*Additional comments on Access Control Rules such as challenges, expectations, etc.*	support for collaboration vs security "spending time in maintaining the policies is a pain" strict access control policies prevent resource misutilization challenges with manual access granting process access control rules should not depend on int/ext member security experts would be consulted to set up access control observed discrepancy in access permissions across groups access control rules could be useful rule system must be user-friendly and simple access control rules have to be clear and evident reasoning behind the rules must be explained new systems should not be difficult to learn "project manager will be the best person to control access" "all close" and "on-deman open" rule for data sharing is useful

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
			difficult to manage/maintain rules in long-term projects
			misuse can be avoided by having rules
			no help from university to introduce rules in private resources
			no time to have rules for private resources
			different tools can be useful for access control (usability)
	OVERALL SATISFACTION	*A binary quantitative data to tag participant's overall experience*	satisfactory
			non-satisfactory
	CHALLENGES MENTIONED BY RESEARCHERS	*Any negtive opinions, feedback, or challeneges shared by researcher participants regarding the current state of RC and possible reasons why their experience with RC is not satisfactory*	default privileges in rc are not setup ideally
			anyone can access others' data in current RC infrastructure
			understanding what levels of privileges a user has is difficult
			rc managed by very few sysadmins
			lack of freedom to manage access in private partitions
			concerns with security when sharing RC with ext collabs
			using rc requires specialized trainings
			in-house resources owned to overcome RC shortcomings
			personal powerful devices over RC
			huge data sharing difficult through university tools
			cluster based data sharing difficult
			issues with RC support team
			submitting jobs is difficult for newcomers
			queue system is frustrating
			permission mgnt is a friction point
			lack of standardization in cluster configs hinder research
			"it could be worse"
			limited resources for the needs
			accidental data deletion from scratch (group-level permissions)
			permissions request for time crucial privileges hamper progress
			IT people do not always know how things work
			unable to update permissions on one's account's data
			group level access can be overwhelming to students
			delayed access revocation unfair for current students
			RC trainings are not practical/beginner-friendly

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
			self-reflection and accountability by admin(s)
			RC help is not enough; users have to figure out stuff
OPINIONS ON CURRENT RC SITUATION	CHALLENGES MENTIONED BY ADMINS	*Any negtive opinions, feedback, or challeneges shared by admin participants regarding the current state of RC and possible reasons why their experience with RC is not satisfactory*	changing mindsets is difficult
			getting everyone to agree policies is challenging
			differing understanding of faculty about scheduler's working
			power struggles and interpersonal factors hinder change
			fitting everyone's need in one big scheduler is not easy
			lack of access policy comprehension/translation
			"research software is very difficult"
			Unix authentication mechanisms are arcane and complicated
			enforcing fine-grained monitoring is challenging
			challenges with fewer admins
			lack of single standard level of access control
			customization vs generalizability
			difficulty picking the right solution
			integrating solutions with existing systems is challenging
			ACLs are not well supported across HPC resources
			unaware if tailored level of access can be achieved using slurm
			keeping up with fundamental paradigm shifts is challenging
			auditing list of active users is challenging
			"it's almost like a Frankenstein"
			rc infrastructure lacks scalability and extensibility
			diverse data policies across organizations is an issue
			configuring Slurm can be complicated
			accessing external resources from RC is difficult
			lowering barrier to entry to RC is critical - skill vs resource
			paradox: performance vs security in HPC/RC (+)
			economic constraints play a major role in RC infrastructure
			institutional reqs or team capabilities influence decisions
			admins have to remind PIs for active user accoutns info
			poor administrative documentation is a barrier to new admins
			documentation is not always updated

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
			running out of resources
			limitations of old infrastructure
	POSITIVE	*Any positive opinions, feedback, or experiences shared by participants regarding the current state of RC and possible reasons why their overall experience with RC is satisfactory*	satisfied with the documentation and the ticketing system
			full administrative access allow admins to help anybody
			RC support for collaboration has been seamless
			satisfied with current privilege management for files
			appreciate that RC doesnot delegate permissions of backups
			huge reliance on RC infrastructure
			RC team is very approachable and helping
			RC educational workshops/trainings are useful
			satisfied with current RC access mgnt implementation
			currently satisfied with RC infrastructure
			reliant on RC because part of university
			resource allocation prioritizes contributors (private)
			more funding could be beneficial to improve RC
	NON-TECHNICAL WISHES	*Desires, suggestions, or ideas pertaining to non-technical aspects of the RC infrastructure, including procedures, and organizational policies, and more.*	more domain experts to support domain-specific research
			lowering the barrier - people intensive vs resource intensive
			more human resources (admins)
			larger RC team for outreach and education
			user HPC community to promote RC Best Practices
			forum for user FAQs (bringing researchers together)
			more education on globus/file permissions
			more tutorials/education to make efficient use of RC
			collecting user feedback to improve administrative work
			better understanding of the research going on in RC
			more balance of responsibilities between admin and user
			improved administrative documentation and knowledge management
			structured onboarding process and training for admins
			homogeneity over heterogeneity in RC systems
			collaboration with different PI's to expand RC systems
			more flexible authentication for RC
			better support for accessing external resources with RC

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
ADMINISTRATIVE WISHES	TECHNICAL AND IMPLEMENTATION WISHES	*Desires, suggestions, or ideas related to the enhancement or implementation of specific technical features, functionalities, tools, or systems within RC*	increasing usability to use RC lowers the barrier to entry better ux support for interactive tools within RC better and simpler software installation process more resources would always be great more proactive measures for vulnerability scanning in VMs better integration w/ university records about active students setting up system to grant access to ext collabs having systems that present RC usage information to users
	TRACKING AND MANAGEMENT WISHES	*Desires, suggestions, or ideas aimed at improving the tracking, monitoring, and management of access privileges and shared resources among users*	a central dashboard (platform) for easy access mgnt delegating student account management to PIs total automation for user account creation/deletion is wished automated double approval process for new accounts verification of user sponsors or PIs by admins Coldfront can help PIs audit access permissions desired access control policies from admins perspective access control for data throughout the project lifecycle a single standard level of access control access control must adapt well with changes from ad-hoc to more structure collboration with researchers better automated ways for file system management simplifying access management improve on accounting information
	NON-TECHNICAL WISHES	*Desires, suggestions, or ideas pertaining to non-technical aspects of the RC infrastructure, including procedures, and organizational policies, and more.*	more targeted scenario-based tutorials for efficient use of RC training for admins on hpc and os internals tutorials to make efficient use of RC training on using containers and dockers could benefit useful to have guides for RC use cases; better education RC team should be properly trained user consultation/input is crucial in decision-making (RC) student representative to advocate for users can be appointed more outreach to promote RC is needed
			a karma system for job priority

THEME NAMES	CATEGORY NAMES	CATEGORY DESCRIPTION	CODE NAME
RESEARCHER WISHES	TECHNICAL AND IMPLEMENTATION WISHES	*Desires, suggestions, or ideas related to the enhancement or implementation of specific technical features, functionalities, tools, or systems within RC*	external collab RC resource more flexible ability to install things is wished better visualization of resources and their status needed more granular resource allocation restrict resource per user more gpus are wished high performance, high memory nodes are wished systematic way to cleanup legacy data confidentiality and integrity for remote data sharing the cluster could benefit with container implementation read-only mount to access large data sets more secure way of data sharing automated expiry of guest access university RCs can do what national funding agency RC is doing
	TRACKING AND MANAGEMENT WISHES	*Desires, suggestions, or ideas aimed at improving the tracking, monitoring, and management of access privileges and shared resources among users*	better transperancy for privileges more granular privilege configuration for undergrad and grads better access configuration for external collaborators examples of existing user/resource access management tools new group between root & member is wished to manage access RC infrastructure can implement Globus like permissions system allow PIs to manage permissions for their students regular check-in system to track access permissions robust RC and file systems for better access mgnt better job priority rules to prevent misuse (access fairness)
	SHARING MECHANISMS WISHES	*Desires, suggestions, or ideas aimed at improving the methods and systems for sharing resources among users within RC*	better ways to mount shared storage device on rc temporary time-based access to shared data for ext collabs group level permissions and access is wished within team allow ext collabs to access internal resources wish to share university cluster data with collaborators