General Information

1. General Information

1.Project Title

Factors in Machine-Level Policy Sharing within the integrated Rule-Oriented Data System (iRODS)

2.**Brief Summary**. Provide a brief non-technical description of the study, which will be used in IRB documentation as a description of the study. Typical summaries are 50-100 words. Please reply to each item below, retaining the subheading labels already in place, so that reviewers can readily identify the content.

Purpose: To empirically examine the motivating and discouraging factors for machine-level policy sharing among iRODS users and partners.

Participants: iRODS users and partners.

Procedures (methods): We will begin the study with a focus group at an upcoming iRODS User Meeting. We will analyze the focus group results and develop a questionnaire. We will examine users' written policies for themes, plus the core.irb files of iRODS users. We will examine what is written vs. what is actually implemented in order to determine any discrepancies. We will analyze these results, develop hypotheses, relate these hypotheses to existing theory and create a model. We will test the model by statistical analysis, to determine the validity and the strengths of the relationships. Pending the results we will conduct interviews and/or a follow-up questionnaire, and then more analysis.

3.Is this submission similar to or related to an application already approved by a UNC-Chapel Hill IRB?

No

2. Project Personnel

1.List all project personnel beginning with principal investigator, followed by all co-investigators, faculty advisor and anyone else who has contact with subjects or identifiable data from subjects. The table below will access campus directory information; if you do not find your name, your directory listing may need to be updated.

Last Name	First Name	Department Name	Role
Moore	Reagan	School of Info & Libr Science	Other
Ward	Jewel	School of Info & Libr Science	Principal Investigator
Tibbo	Helen	School of Info & Libr Science	Faculty Advisor
Whittington	Mary	RENCI	Other
Conway	Mike	DICE Center	Other
De Torcy	Antoine-Cecil	School of Info & Libr Science	Other
Rajasekar	Arcot	School of Info & Libr Science	Other

NOTE: The IRB database will link automatically to UNC Human Research Ethics Training database and the UNC Conflict of Interest (COI) database. IRB will communicate with the personnel listed above or the PI if further documentation is required.

2.Center, institute, or department in which research is based (Administering Department) if other than department(s) listed above:

Department	Renaissance Computing Inst

3. Funding Sources

1.Is this project funded (or proposed to be funded) by a contract or grant that is sponsored by an organization external to UNC-Chapel HIll?

No

2.Is this study funded by UNC-CH (e.g., department funds, internal pilot grants, trust accounts)?

Yes

Internal UNC Chapel Hill funding

Department Name	Account Number
School Of Information And Library Science	

3.Is this research classified (e.g. requires governmental security clearance)?

No

4.Is there a master protocol, grant application, or other proposal supporting this submission (check all that apply)?

- ✖ Grant Proposal
- ✖ Industry Sponsor Master Protocol
- ✖ Student Dissertation or Thesis Proposal
- ✖ Investigator Initiated Master Protocol
- ✖ Other Study Protocol

4. Screening Questions

1.Is your project Human Subjects Research (HSR) that requires IRB review and approval (e.g., clinical trial, identifiable survey)? Select "No" if either you are unsure or you would like documentation that your project does not require IRB approval.

Yes

The following questions will help build the remaining components of your IRB application.

A. Will you be using information (data, records or human biological specimens) that is currently existing or will be collected in the future for purposes other than this proposed research (e.g., medical records, ongoing collection of specimens for a tissue repository)?

Yes

B. Will you be collecting data about a living individual through direct intervention or interaction with that individual? This includes any contact with subjects including surveys, questionnaires, interviews, focus groups, observations, treatment intervention(s), etc.

No

C. Will subjects be studied in the Clinical and Translational Research Center (CTRC, previously known as the GCRC) or is the CTRC involved in any other way with the study? If yes, this application will be reviewed by the CTRC and additional data will be collected.)

No

D. Does this study involve UNC-CH cancer patients? (If yes, this application may require additional review by the Oncology Protocol Review Committee.)

No

E. Are any personnel, organizations, entities, facilities or locations in addition to UNC-Chapel Hill involved in this research (e.g., is this a multi-site study or does it otherwise involve locations outside UNC-CH, including foreign locations)?

Yes

5. Multi-site Study Information

1.Will this study be conducted in locations outside the United States?

Yes

Will your research project involve the Galapagos Islands, Ecuador?

No

If yes, your application will be reviewed by the UNC Center for Galapagos Studies. This Center will be included in routing for approvals after you submit.

Are any of the countries on the U.S. State Department Travel Warning List? See list found at http://www.trave http://www.travel.state.gov (look for "Travel Warnings"). See also the University policy at http://www.unc.edu/campus/policies/Travel%20Warning%20Policy.pdf

No

2.Is UNC-CH the sponsor, lead coordinating site or otherwise responsible for this study being conducted at sites outside of UNC-CH?

Yes

Complete the following information for each site outside of UNC-CH:

Name	City	State	Country	Status of IRB approval
N/A	N/A	N/A	N/A	N/A

3.Describe the role of UNC Chapel Hill and UNC Chapel Hill investigator(s) in this study.

The study investigator will be conducting a focus group with users and partners of DICE and iRODS during an annual user group meeting, which is held at UNC-CH. The users and partners attending the conference live world-wide. Later, the study investigator will be sending links to questionnaires via email and/or interviewing these individuals. The questionnaires will be online via a tool such as Qualtrics. The interviews will be conducted online or via telephone. Some interviews may be conducted in person, pending the location of the iRODS user. The study investigator will be at UNC; participants are located around the world.

Researchers are reminded that additional approvals may be needed from relevant "gatekeepers" to access subject.

Part A. Questions Common to All Studies

A.1. Background and Rationale

1.Provide a summary of the background and rationale for this study (i.e, why is the study needed?). If a complete background and literature review are in an accompanying grant application or other type of proposal, only provide a brief summary here. If there is no proposal, provide a more extensive background and literature review, including references.

Librarians, archivists, ILS researchers and computer scientists have made substantial efforts over the past two decades to define the requirements for implementing a trustworthy digital preservation repository (Lynch, 1994; Waters & Garrett, 1996; Bearman & Trant, 1998; Council on Library and Information Resources, 2000; Research Libraries Group, 2002, 2005;). Their efforts have culminated in the creation of general guidelines and policies, such as the Reference Model for an Open Archival Information System (OAIS) (Consultative Committee for Space Data Systems, 2002), Trustworthy Repositories Audit & Certification (TRAC) (Online Computer Library Center & Center for Research Libraries, 2007), ISO MOIMS-RAC (Consultative Committee for Space Data Systems, 2009), or self-audits such as DRAMBORA (Digital Curation Centre & Digital Preservation Europe, 2007). These practitioners and researchers have sought to establish and define the characteristics of a trustworthy repository, and then provide the mechanisms to verify that a repository actually does fulfill the established criteria for trustworthiness (Consultative Committee for Space Data Systems, 2009).

One way to audit and verify that the administrators of a digital repository meet and enforce these standards for trustworthiness is to automate the policies within the system itself. For example, the creators of the middleware data grid system, the iRule Oriented Data System (iRODS), designed it so that repository managers and administrators could implement archival policies at the machine-level as "rules" (Rajasekar, 2006). This system built on previous data grid research and experience (Moore & Merzky, 2003; Moore, 2006) with the Storage Resource Broker (SRB) (Moore, 2004; Moore, 2005). One research project, called PLEDGE (MacKenzie & Moore, 2006), bridged the gap between written policies that "should" be implemented at the machine-level by identifying which TRAC guidelines "could" be implemented at the machine-level within SRB.

While the primary idea behind establishing trusted repository audit mechanisms is to ensure that a repository actually meets and enforces archival standards (Consultative Committee for Space Data Systems, 2009), the other benefit of using standards is reduced costs (Science and Technology Council, 2007). By creating a standard set of machine-level automated repository policies based on archival standards, the amount of human intervention needed in the digital preservation process will be reduced because these policies may then be shared among and between communities. This will streamline the digital archive process and potentially reduce long-term costs, thus aiding the longevity of the archive, as well as the integrity of the metadata and digital content it contains. The fewer resources that archivists require to maintain an archive, the more likely an institution or organization is to maintain it for the indefinite long-term.

In theory, the process of taking written policies and implementing them as machine-actionable rules should be fairly simple. In reality, the process of creating rules out of written guidelines has been slow to catch on within a variety of communities that use preservation repositories. The assumption behind the establishment of standards such as ISO MOIMS-RAC is that these policies will be implemented within repositories. Other practitioners and researchers have assumed that the most streamlined way to implement these standards is at the machine-level (Hunter & Choudhery, 2005). These are, after all, digital repositories. The creators of iRODS assumed that communities would coalesce and create rules and micro-services (Moore, Rajasekar & Wan, 2009). While this has happened in some communities, this has not happened in others. The possible impediments include the learning curve for new technologies, whether the community has quantified their policies, whether the community has infrastructure that supports policy-based management for digital data, and whether the standard for assessment criteria represents a viable set of policies. If the theory behind almost two decades of work by librarians, archivists, ILS researchers and computer scientists working on the digital preservation problem is that policies will be implemented at the machine-level, but in reality practitioners aren't doing this, then some of the basic theories of digital preservation for the past 15 years are at stake.

If practitioners aren't accepting these theories as "best practices" and implementing them, is it because the theories are wrong, or are they untenable? The current hypothesis within the iRODS group is that community members who manage preservation repositories have policies that are unique to their community, but they find it challenging to implement written policies at a machine-level, and that only a few communities have implemented more than a core (provided) set of policies at the machine-level. The purpose of this study is to empirically study the motivating and discouraging factors for machine-level policy sharing within the iRODS community.

Most of the related literature examines the creation of the policies themselves. These authors have not examined the actual implementation of the policies within a preservation system; therefore, there is little to no previous direct work in this specific area.

References

Bearman, D. & Trant, J. (1998, June). Authenticity of digital resources: towards a statement of requirements in the research process. D-Lib Magazine. Retrieved April 14, 2009, from http://www.dlib.org/dlib/june98/06bearman.html

Consultative Committee for Space Data Systems. (2002). Reference model for an Open Archival Information System (OAIS) (CCSDS 650.0-B-1). Washington, DC: National Aeronautics and Space Administration (NASA). Retrieved April 3, 2007, from http://nost.gsfc.nasa.gov/isoas/

Consultative Committee for Space Data Systems. (2009). Audit and Certification of Trustworthy Digital Repositories (CCSDS 652.0-R-1). Washington, DC: National Aeronautics and Space Administration (NASA). Retrieved July 27, 2010, from http://wiki.digitalrepositoryauditandcertification.org/pub/Main/WebHome/652x0r1candidate-update-typoscorrected.doc

Council on Library and Information Resources. (2000). Authenticity in a digital environment. Washington, DC: Council on Library and Information Resources. Retrieved September 11, 2007, from http://www.clir.org/pubs/reports/pub92/pub92.PDF

Digital Curation Centre & Digital Preservation Europe. (2007). DCC and DPE digital repository audit method based on risk assessment (DRAMBORA). Retrieved August 1, 2007, from http://www.repositoryaudit.eu/download

Hunter, J. & Choudhury, S. (2005). Semi-automated preservation and archival of scientific data using semantic grid services. In Proceedings of the IEEE International Symposium on of Cluster Computing and the Grid, May 9-12-2005, Cardiff, UK.

Lynch, C. A. (1994). The integrity of digital information: mechanics and definitional issues. Journal of the American Society for Information Science, 45(10), 737-744.

MacKenzie, S. & Moore, R. (2006). Digital Archive Policies and Trusted Digital Repositories. Paper presented at the 2nd International Digital Curation Conference, November 21 - 22, 2006, Glasgow, Scotland. Retrieved November 2, 2009, from http://pledge.mit.edu/images/6/6f/Smith-Moore-DCC-Nov-2006.pdf

Moore, R. (2004). Evolution of data grid concepts. Paper presented at the workshop on “data” at the 10th Global Grid Forum, Berlin, Germany, March 9-13, 2004. Retrieved March 23, 2009, from http://www.npaci.edu/DICE/Pubs/Grid-evolution.doc

Moore, R. (2005). Persistent collections. In S.H. Kostow & S. Subramaniam (Eds.), Databasing the brain: from data to knowledge (neuroinformatics) (pp. 69-82). Hoboken, NJ: John Wiley and Sons.

Moore, R. (2006). Building preservation environments with data grid technology. American Archivist, 69(1), 139-158.

Moore, R. & Merzky, A. (2003). Persistent archive concepts. Paper presented at the 7th Global Grid Forum, Tokyo, Japan, March 4-7, 2003. Retrieved March 4, 2009, from http://www.npaci.edu/DICE/Pubs/Data-PAWG-PA.doc

Moore, R., Rajasekar, A., & Wan, M. (2005). Data grids, digital libraries and persistent archives: an integrated approach to publishing, sharing and archiving data. Proceedings of the IEEE (Special Issue on Grid Computing), 93(3). Retrieved July 24, 2007, from http://ieeexplore.ieee.org/iel5/5/30407/01398012.pdf?arnumber=1398012

Online Computer Library Center, Inc. & Center for Research Libraries. (2007). Trustworthy repositories audit & certification: criteria and checklist version 1.0. Dublin, OH & Chicago, IL: OCLC & CRL. Retrieved July 27, 2010, from http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Rajasekar, A., Wan, M., Moore, R., & Schroeder, W. (2006). A prototype rule-based distributed data management system. Paper presented at a workshop on “next generation distributed data management" at the High Performance Distributed Computing Conference, June 19-23, 2006, Paris, France.

Research Libraries Group. (2002). Trusted digital repositories: attributes and responsibilities an RLG-OCLC report. Mountain View, CA: Research Libraries Group. Retrieved September 11, 2007, from http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf

Research Libraries Group. (2005). An audit checklist for the certification of trusted digital repositories, draft for public comment. Mountain View, CA: Research Libraries Group. Retrieved April 14, 2009, from http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2416.pdf

Science and Technology Council. (2007). The digital dilemma strategic issues in archiving and accessing digital motion picture materials. The Science and Technology Council of the Academy of Motion Picture Arts and Sciences. Hollywood, CA: Academy of Motion Picture Arts and Sciences.

Waters, D. and Garrett, J. (1996). Preserving Digital Information. Report of the Task Force on Archiving of Digital Information. Washington, DC: CLIR, May 1996.

2.State the research question(s) (i.e., specific study aims and/or hypotheses).

(1) Are iRODS users' policies written, unwritten, or non-existent?
(2) What are the motivating & discouraging factors for machine-level policy sharing within the iRODS community?
(3) What are the benefits and costs of machine-level policy sharing within the iRODS community?
(4) Can these policies (written and unwritten) be grouped by domain? Or, do all organizations share a common base policy set? And, if so, what does that base policy set contain? If there are differences between domains, how/where are they different?
(5) Can these policies (written and unwritten) be grouped by function? If so, how? If not, why?
(6) Can iRODS base policies (core.irb) be categorized by function and/or domain?

A.2. Subjects

1.Number of subjects across all sites (provide exact number or range):

20-30 (estimate)

2.Number of subjects at UNC-CH (provide exact number or range):

3-6

3.Do you have specific plans to enroll subjects from these vulnerable or select populations:

✗ Children (under the age of majority for their location)

Note that you will be asked to provide age ranges for children in the Consent Process section.

✗ Non-English-speaking	
✗ Patients (i.e., have a specific disease, disorder or condition regardless of where they receive their healthcare)	
✗ Prisoners, others involuntarily detained or incarcerated, or parolees	
✗ Decisionally impaired	
✗ Pregnant women	
✗ HIV positive individuals	
✗ UNC-CH Students	
Some research involving students may be eligible for waiver of parental permission (e.g., using departmental participant pools) See SOP 32.9.1	
✓ UNC-CH Employees	
✗ People who are likely to be involved in abusive relationships, either as perpetrator or victim (Link to guidance = appendix I from SOPs)	

4.Describe any mechanism that you plan to use to confirm status in one or more of the above groups (e.g., pregnancy, psychological or HIV testing)

I will ask participants in the focus group or questionnaire for information on where they work and their job title via a sign up sheet (focus group) or as part of the questionnaire.

5.If any of the above vulnerable populations are checked, please describe your plans to provide additional protections for these subjects

I will keep the data password protected on my computer, or in a locked drawer in my desk. I will anonymize the data contributors' name, etc.

6.Age range of subjects:

Minimum age of subject enrolled	18
Unit of time	years
Maximum age of subject enrolled	99
» If no maximum age limit, indicate 99	
Unit of time	years

A.3. Inclusion/exclusion criteria

1.List required characteristics of potential subjects (i.e., inclusion and exclusion criteria). If not covered, list also characteristics that would preclude their involvement.

Subjects must be data grid managers using iRODS. DICE employees at UNC-CH and UCSD will be excluded from participation.

2.Justify any exclusion based on race, gender or ethnicity

N/A

3.Will pregnant women or women who become pregnant be excluded or withdrawn?

No

A.4. Study design, methods and procedures

1.Describe the study design. List and describe study procedures, including a sequential description of what subjects will be asked to do, when relevant.

The study is mixed methods (qualitative and quantitative). It is a case study in that only users of the iRODS system will be in the sample population studies, but is is primarily a content analysis, in that I will be analyzing focus group responses, survey responses, written policies (human language) and machine-level policies (computer code).

1. We run a focus group during an annual meeting. Study participants will be asked a series of non-leading questions related to the research questions. Their responses will be recorded.
2. We code and analyze the responses from the focus group.
3. Based on the results of the focus group, we create an online survey on machine-level policy implementation and submit to participants. Participants will be asked to fill out a series of open answer and closed answer responses.
4. We code and analyze the responses to the survey.
5. We examine, code, and analyze both written policies (human language) and machine-actionable (computer code).
6. Taking existing theory in the literature and the results of 1-5 above, we create one or more hypothesis regarding machine-level policy sharing.
7. Based on the hypothesis created in 6 above, we create a model of machine-level policy sharing that describes the factors motivating and discouraging machine-level policy sharing.
8. We test the model using statistical analysis to determine the strengths of the relationships between the various factors.
9. We determine whether or not the model matches existing theory on code sharing in open source communities.
10. Based on the model, if needed we submit a follow-up questionnaire to study participants or conduct individual interviews. Participants will be asked to respond to a series of questions in an online survey or via telephone or online interview.
11. Analyze, write up and publish any relevant results to the wider research community.

2.If not already described above, if subjects are assigned or randomized to study "arms" or groups, describe how they are assigned.

N/A

3.Describe any follow up procedures.

Based on the model, if needed we will submit a follow-up questionnaire to study participants or conduct individual interviews. In that instance, participants will be asked to respond to a series of questions in an online survey or via telephone or online interview. I will revise this IRB at that time to include any follow up surveys.

4.Duration of subject's participation

The focus group will last approximately 2 hours. The time required for the online surveys will vary according to how much information the participant provides. I estimate the surveys will take anywhere from 15 minutes to 2 hours, depending on the level of response.

5.Duration of entire study (include start and stop dates, where known).

February 2011 to August 2012.

Your responses to the next set of questions will help determine what further questions you will be asked in the following sections.

6.Will this study use of any of the following methods?

✔ Audiotaping
✘ Videotaping or filming
✘ Behavioral observation - (e.g., Participant, naturalistic, experimental, and other observational methods typically used in social science research)
✘ Pencil and paper questionnaires or surveys
✔ Electronic questionnaires or surveys

- ✔ Telephone questionnaires or surveys
- ✘ Interview questionnaires or surveys
- ✘ Other questionnaires or surveys
- ✔ Focus groups
- ✘ Diaries or journals
- ✘ Photovoice
- ✘ Still photography

7.Will you be using any methods commonly used in biomedical or clinical research (e.g., blood draws or other clinical lab tests, biological monitoring, physical examination)?

No

8.If there are procedures or data collection methods that require specialized training, describe who (role/qualifications) will be involved and how they will be trained.

N/A

9.Are there cultural issues, concerns or implications for the methods to be used with this study population?

No

A.5. Benefits to subjects and/or society

1.Describe the benefit to society based on scientific knowledge to be gained

This study may aid in the long-term preservation of digital materials.

While the primary idea behind establishing trusted repository audit mechanisms is to ensure that a repository actually meets and enforces archival standards (Consultative Committee for Space Data Systems, 2009), the other benefit of using standards is reduced costs (Science and Technology Council, 2007). By creating a standard set of machine-level automated repository policies based on archival standards, the amount of human intervention needed in the digital preservation process will be reduced because these policies may then be shared among and between communities. This will streamline the digital archive process and potentially reduce long-term costs, thus aiding the longevity of the archive, as well as the integrity of the metadata and digital content it contains. The fewer resources that archivists require to maintain an archive, the more likely an institution or organization is to maintain it for the indefinite long-term.

References

Consultative Committee for Space Data Systems. (2009). Audit and Certification of Trustworthy Digital Repositories (CCSDS 652.0-R-1). Washington, DC: National Aeronautics and Space Administration (NASA). Retrieved July 27, 2010, from http://wiki.digitalrepositoryauditandcertification.org/pub/Main/WebHome/652x0r1candidate-update-typoscorrected.doc

Science and Technology Council. (2007). The digital dilemma strategic issues in archiving and accessing digital motion picture materials. The Science and Technology Council of the Academy of Motion Picture Arts and Sciences. Hollywood, CA: Academy of Motion Picture Arts and Sciences.

2.Does this study have the potential for direct benefit to individual subjects in this study?

Yes

Consider the nature, magnitude, and likelihood of any direct benefit to subjects. If there is no direct benefit to the individual subject, say so here and in the consent form, if there is a consent form. Do not cite monetary payment or other compensation as a benefit.

Explain

The direct benefit to the participants is the ability to create a structure (social and technical) that will aid in the implementation of policies at the machine-level. At a high level, this can contribute to the preservation and maintenance of the digital materials in the repository. On a more day-to-day level, if may be able to create turnkey solutions for grid administrators. For example, if grid manager X manages a physics data repository, he or she needs Policy Set A. If grid

manager Y manages a cultural heritage image repository, he or she needs Policy Set B.

3.Are there plans to communicate the results of the research back to the subjects?

Yes

If yes, describe

Via a summary of results in project reports, published results, or online via a link to all results, with any identifying information removed. If email is used, I will email participants individually.

A.6. Risks and measures to minimize risks

For each of the following categories of risk you will be asked to describe any items checked and what will be done to minimize the risks. Where possible, describe the likelihood of the risks occurring, using the following terms:

- ***Very Common*** *(approximate incidence > 50%)*
- ***Common*** *(approximate incidence > 25%)*
- ***Likely*** *(approximate incidence of 10-25%)*
- ***Infrequent*** *(approximate incidence of 1-10%)*
- ***Rare*** *(approximate incidence < 1%)*

1.Psychological

✗	Emotional distress
✓	Embarrassment
✗	Consequences of breach of confidentiality
✗	Other

Describe any items checked above and what will be done to minimize these risks

Rare. To minimize this risk, data will be anonymized, including participant and institution names.

2.Social

✓	Loss of reputation or standing within the community
✗	Harms to a larger group or community beyond the subjects of the study (e.g., stigmatization)
✓	Consequences of breach of confidentiality
✗	Other

Describe any items checked above and what will be done to minimize these risks

Rare. To minimize these risks, data will be anonymized, including participant and institution names.

3.Economic

✗	Loss of income
✓	Loss of employment or insurability
✓	Loss of professional standing or reputation
✓	Loss of standing within the community
✓	Consequences of breach of confidentiality
✗	Other

Describe any items checked above and what will be done to minimize these risks

Rare. To minimize these risks, data will be anonymized, including participant and institution names.

4.Legal

✗	Disclosure of illegal activity
✗	Disclosure of negligence
✓	Consequences of breach of confidentiality
✗	Other

Describe any items checked above and what will be done to minimize these risks

Rare. To minimize these risks, data will be anonymized, including participant and institution names.

5.Physical

✗	Medication side effects
✗	Pain
✗	Discomfort
✗	Injury
✗	To a nursing child or a fetus (either through mother or father)

6.Unless already addressed above, describe procedures for referring subjects who are found, during the course of this study, to be in need of medical follow-up or psychological counseling

N/A.

7.Will there be additional follow up for subjects who become pregnant or father a child while enrolled in this study?

No

8.Are there plans to follow a female partner of a male subject if she becomes pregnant while he is enrolled in this study?

No

A.7. Data and safety monitoring

1.Describe the plan for monitoring the data to ensure the safety of participants. These plans could range from the investigator monitoring subject data for any safety concerns to a sponsor-based data and safety monitoring board or committee (DSMB, DSMC, DMC), depending on the study.

The data will be password protected on the PIs computer. The PIs laptop requires a password to login to the laptop. Access to any of the passwords stored in the PIs password storage (e.g., "keychain") requires a separate password. Any paper containing participant information will be in a locked filing cabinet. Backups of the PIs computer are made daily; only the PI has access to the backups.

2.If not already addressed above, describe the plans for aggregate review of unanticipated problems (including but not limited to adverse events) across all sites, in order to monitor subject safety.

The PI will maintain contact with the participants, and ask each to advise her if any complications arise.

3.What are the criteria that will be used to withdraw an individual subject from this study or halt the research intervention (e.g., abnormal lab tests, allergic reactions, failure or inability to comply with study procedures, etc.)?

The participant will be withdrawn from the study if they change jobs -- i.e., are no longer a grid manager using iRODS. A participant may be withdrawn if she or he does not follow instructions.

4.Are there criteria that will be used to stop the entire study prematurely (e.g., safety, efficacy, unexpected adverse events, inability to recruit sufficient number of subjects, etc.)?

Yes

Please explain

If the PI cannot recruit participants she will stop the study.

5.Will this study involve a data and safety monitoring board or committee?

No

A.8. Data analysis

1.Describe the analytical methods to be used (qualitative or quantitative)

Mixed methods, qualitative and quantitative. This is a case study in that only iRODS will be studied, but the primary method is a content analysis. The quantitative component comes in during the statistical analysis of the results of the data generated from the content analysis, as well as the testing of the model (the strength of the relationship between factors motivating and discouraging the sharing of machine-level policies). I cannot provide the names of particular statistical tests at this time, as I am not sure what the data will require. I have not yet gathered the data.

2.Explain how the sample size is sufficient to achieve the study aims. This might include a formal power calculation or an explanation of why a small sample is sufficient (e.g., qualitative research, pilot studies)

For the focus group, I'll aim for the standard 6-12 participants.

The sample population will be determined from the number of users registered on the iRODS chat mailing list. Currently, I estimate there are about 355 non-DICE members of the chat list.

Assuming a population of 355, with a Confidence Interval of 15 and a Confidence Level of 15 I'll need approximately 38 people (grid managers) to respond to the survey(s).

A.9. Identifiers

1.Check all of the following identifiers you will be receiving. This does not apply to information on consent forms:*

✔	Names
✔	Telephone numbers
✘	Any elements of dates (other than year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death. For ages over 89: all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 and older
✔	Any geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code and their equivalent geocodes (e.g. GPS coordinates), except for the initial three digits of a zip code
✘	Fax numbers
✔	Electronic mail addresses
✘	Social security numbers
✘	Medical record numbers
✘	Health plan beneficiary numbers
✘	Account numbers
✘	Certificate/license numbers
✘	Vehicle identifiers and serial numbers (VIN), including license plate numbers
✘	Device identifiers and serial numbers (e.g., implanted medical device)
✘	Web universal resource locators (URLs)

- ✗ Internet protocol (IP) address numbers
- ✗ Biometric identifiers, including finger and voice prints
- ✗ Full face photographic images and any comparable images
- ✗ Any other unique identifying number, code, or characteristic, other than dummy identifiers that are not derived from actual identifiers and for which the re-identification key is maintained by the health care provider and not disclosed to the researcher

2.For any identifiers checked, how will these identifiers be stored in relationship to the research data?

- ✓ with the research data (i.e., in the same data set and/or physical location)
- ✗ separate from the research data (i.e., coded with a linkage file stored in a different physical location)

3.If collecting Social Security Numbers, will the subjects' Social Security Number (SSN) be collected for use as a unique indentifier for study tracking purposes for national registry or database?

No

A.10. Confidentiality of the data

1.Describe procedures for maintaining confidentiality of the data you will collect or will receive (e.g., coding, anonymous responses, use of pseudonyms, etc.).

I will use coding and pseudonyms to maintain confidentiality of data.

2.Describe how data will be transmitted among research team (i.e., personnel listed on this application).

The data will not be transmitted. It will reside on the PIs computer and will be backed up on a daily basis. Only the PI has access to the back up files.

3.Are you collecting sensitive information such as sexual behavior, HIV status, recreational drug use, illegal behaviors, child/physical abuse, immigration status, etc?

No

4.Do you plan to obtain a federal Certificate of Confidentiality for this study?

No

5.If relevant, discuss the potential for deductive disclosure (i.e., directly identifying subjects from a combination of indirect IDs).

It is possible, but highly unlikely. Someone would have to have access to the iRODS chat membership list, know what institutions use iRODS, their location, and the grid managers' name.

6.Will any of the groupings or subgroupings used in analysis be small enough to allow individuals to be identified?

No

A.11. Data sharing and transmission

1.Check all of the following who will receive identifiable data (contains any of the 18 identifiers listed above) outside the immediate research team (i.e., not listed as personnel on this application)? *

- ✓ No one
- ✗ Coordinating Center
- ✗ Statisticians
- ✗ Consultants

- ✗ Other researchers
- ✗ Registries
- ✗ Sponsors
- ✗ External labs for additional testing
- ✗ Journals
- ✗ Publicly available dataset
- ✗ Other

2.For any recipients checked above, explain the confidentiality measures to be taken

No Answer Provided

A.12. Post-study disposition of identifiable data or human biological materials

1.Describe your plans for disposition of data or human biological specimens that are identifiable in any way (directly or via indirect codes) once the study has ended. If you plan to destroy linkage codes or identifiers, describe how this will be done.

Once the study has been completed, I will de-identify the data and make it publicly available. I will keep the identifiable file information on my computer, password protected, or in a locked filing cabinet.

Part C. Existing Data, Records, Specimens

C.1. Data Sources

1.What existing records, data or human biological specimens will you be using? (indicate all that apply): *

- ✗ Data already collected from another research study

  Were the investigators for the current application involved in the original collection: --
- ✗ Patient specimens (tissues, blood, serum, surgical discards, etc.)

  Has the clinical purpose for which they were collected been met before removal of any excess? --
- ✗ Data already collected for administrative purposes
- ✗ Medical records in any format, including paper or electronic. This would include MIMS, WebCIS, Carolina Data Warehouse (CDW).

  Be aware that the medical record custodian may also require their own form, e.g., HD-974 if UNC-Health Care System
- ✗ Data coming directly from a health plan, health care clearinghouse, or health care provider?
- ✗ Publicly available data
- ✓ Other

If you have checked any of the above items, provide a description of the data you propose to use, describing the type of data, how they were collected (including consent procedures), and where they currently reside.

Written policies about the digital objects in a repository, and core.irb files. All iRODS installs have a core.irb file in the base software. I will collect the policies and the files with the consent of the grid managers. I cannot access the core.irb file without permission of the grid manager, as it is in the software. Written policies may be available online, but I will only gather those policies if the grid manager provides access to the core.irb files. (There is no point to analyzing any online written policies without also being able to check the core.irb files.)

2.For each of the boxes checked above, do you have permission from the custodians of the data, records or specimens (e.g., pathology dept, tissue bank, original researcher)?

No

Explain

Not yet. As stated, I cannot access the core.irb file myself because it is in the iRODS software. The grid manager has to give it to me.

Do the custodians of the data, records or specimens require a data use agreement?

No

C.2. Coding and Data Use Agreements

1.When you receive these data, records or human biological specimens will they be coded? Coded means identifying information that would enable the research team to readily ascertain the individual's identity has been replaced with a number, letter, symbol, or combination thereof (i.e., a code).

No

Part D. The Consent Process

1. Obtaining informed consent from subjects

The standard consent process is for all subjects to sign a document containing all the elements of informed consent, as specified in the federal regulations. Some or all of the elements of consent, including signatures, may be altered or waived under certain circumstances. If you will be requesting a waiver answer "not applicable" for any of the following questions that will not pertain to this study. You will be asked to provide relevant information in the section below on waivers.

1.Will children under the age of majority in their locale (18 years in NC) be enrolled?

No

2.Will adult subjects be enrolled in your study?

Yes

Explain the process for obtaining consent from the subject or the subject's legally authorized representative, if relevant

I will provide a consent form to the participants in person on paper (for the focus group) or embedded in online survey. For one-on-one interviews by phone, I will ask the participant to complete an online consent form prior to beginning the interview.

3.Will decisionally-impaired subjects be enrolled in your study? (includes unconscious patients, some psychiatric disorders, others who lack the capacity to give consent)

No

4.Are you planning to obtain consent from any Non-English speaking subjects?

No

5.Describe who (name and role) will be obtaining consent or parental permission.

Jewel Ward, the PI, will obtain consent from participants.

6.Describe any steps that will be taken to minimize coercion or undue influence during the consent process. These might include a waiting period between the initial consent discussion and obtaining consent, or obtaining consent by someone other than a person with perceived authority (e.g., professor, employer, treating physician).

For the focus group, no DICE members will be present. For the surveys and interviews, I will assure participants of confidentiality, and that they may withdraw from the study at any time.

7.Has the sponsor of this study provided a model consent form?

No

2. Waiver of written documentation of informed consent

The default is for subjects to sign a written document that contains all the elements of informed consent. Under limited circumstances, the requirement for a signed consent form may be waived by the IRB. For example, this might occur for phone or internet surveys, when a signed consent form is either impractical or unnecessary, or in circumstances where a signed consent form creates a risk for the subject.

1.Are you requesting a waiver of any aspect of written (signed) documentation?

Yes

Choose which of the following consent approaches apply and attach the relevant document: *

- ✔ Full consent form minus the signature lines
- ✘ Information or fact sheet (streamlined unsigned consent form)
- ✔ Online consent form with electronic agreement
- ✔ Consent statement incorporated into a survey itself
- ✘ Verbal consent obtained in person or via the phone
- ✘ Short form (for subjects with limited ability to read full consent form) view instruction
- ✘ Other

Choose which one of the following justifies the waiver of written documentation: *

- ✘ The only record linking the subject and the research would be the consent document and the principal risk would be potential harm resulting from a breach of confidentiality (e.g., study topic is sensitive so that public knowledge of participation could be damaging). Participants should be asked whether they want documentation linking them with the research and the participants' wishes will govern whether they sign the form. Note: This justification cannot be used in FDA-regulated research.
- ✔ The research presents no more than minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context. (e.g., many phone or mail surveys, "man in the street" interviews, etc.).

Explain

The research is about a software system the users' use in their daily jobs. It is not about the users themselves. The information I am seeking relates to a repositories overall policies, not the individuals themselves.

If your request for a waiver of written documentation applies to some but not all of your subject groups and/or consent forms, please describe and justify

No Answer Provided

3. Full or partial waiver of consent

The default is for subjects to give informed consent. A waiver might be requested for research involving only existing data or human biological specimens. More rarely, it might be requested when the research design requires withholding some study details at the outset (e.g., behavioral research involving deception). In limited circumstances, parental permission may be waived. This section should also be completed for a waiver of HIPAA authorization if research involves Protected Health Information (PHI) subject to HIPAA regulation, such as patient records.

1.Are you requesting any of the following:

✘ a waiver of informed consent in its entirety

✗ a waiver or alteration of some of the elements of informed consent

✗ a waiver of HIPAA authorization (If you are accessing patient records for this research, you must also request a waiver of HIPAA authorization)

2.If your request for a waiver applies to some but not all of your subject groups and/or consent forms, please describe and justify

No Answer Provided

3.Does this request for waiver support a study design that involves deception or withholding of information?

No Answer Provided

Consent Forms

	File Name	Document Type
✓	Adult Consent Form	Adult Consent Form
✓	Focus Group Consent	Focus Group Consent
✓	Text for Consent Embedded in Survey	Text for Consent Embedded in Survey
✓	Text for Online Consent Form	Text for Online Consent Form

view consent forms

Attachments

	File Name	Document Type
✓	ElectronicQuestionnaireSurvey-IRBApp.docx Required	Electronic Questionnaire Survey
✓	FocusGroupGuide-IRBApp.docx Required	Focus Group Guide
✓	TelephoneQuestionnaireSurvey-IRBApp.docx Required	Telephone Questionnaire Survey

view attachments

Addenda

Data Security Requirements

view addenda

By certifying below, the Principal Investigator affirms the following:

I will personally conduct or supervise this research study. I will ensure that this study is performed in compliance with all applicable laws, regulations and University policies regarding human subjects research. I will obtain IRB approval before making any changes or additions to the project. I will notify the IRB of any other changes in the information provided in this application. I will provide progress reports to the IRB at least annually, or as requested. I will report promptly to the IRB all unanticipated problems or serious adverse events involving risk to human subjects. I will follow the IRB approved consent process for all subjects. I will ensure that all collaborators, students and employees assisting in this research study are informed about these obligations. All information given in this form is accurate and complete.

This study proposes research that has been determined to include Security Level 2 data security requirements. I agree to accept responsibility for managing these risks appropriately in consultation with departmental and/or campus security personnel. The Data Security Requirements addendum can be reviewed here.

If PI is a Student or Trainee Investigator, the Faculty Advisor also certifies the following:

I accept ultimate responsibility for ensuring that this study complies with all the obligations listed above for the PI.

Certifying Signatures:

Signature: _____ **Date:** _____
Jewel Ward

Signature: _____ **Date:** _____
Helen Tibbo

The expectation is that this approval is being given on behalf of the head of the Department, Division, or Center. If the chair or director is an investigator on this project or otherwise conflicted in approving it, the Vice-Chair or Chair's designee should review it. By approving, you are certifying the following on behalf of your department, division or center:

- This research is appropriate for this Investigator and our department
- The investigator(s) are qualified to conduct the research
- There are adequate resources (including financial, support and facilities) available
- For units that have a local review committee for pre-IRB review, this requirement has been satisfied
- I support this application, and hereby submit it for further review

This study proposes research that has been determined to include Security Level 2 data security requirements. I agree to accept responsibility for managing these risks appropriately in consultation with departmental and/or campus security personnel. The Data Security Requirements addendum can be reviewed here.

If you are approving for other purposes (e.g., CTRC, DSMB, IBC, PRC, RSC, or other review committees), you affirm the following:

- The proposed submission is approved and may be forwarded for IRB review.

This study proposes research that has been determined to include Security Level 2 data security requirements. I agree to accept responsibility for managing these risks appropriately in consultation with departmental and/or campus security personnel. The Data Security Requirements addendum can be reviewed here.

Department Approval Signatures:
By signing in the appropriate space, the Department Chairperson(s) is indicating only that he/she has seen and reviewed this submission

Department: School of Info & Libr Science
Signature: _____ **Date:** _____
Name & Title: _____

Department: Renaissance Computing Inst
Signature: _____ **Date:** _____

Name & Title: _____