**Regulatory Data Environment – Needs and Requirements (ARCHIVED)**

**Notice of Purpose**

The purpose of the Regulated Data Environment (RDE) Needs and Requirements document had originally been intended as to address faculty questions asked in response to the proposal of a centralized RDE.

As additional questions and considerations came about answering that initial inquiry, the document evolved into an outline to address implementation and maintenance processes broken down to the four categories of "Scope", "Cost/Scalability", "Security", and "Support".

Throughout early committee discussions ITS committed resources and leveraged existing cloud infrastructure and personnel capabilities to provide an initial RDE offering for regulated data implementations.

While other documentation and processes have since been developed to help accommodate implementation of an RDE, review of this document is recommended to help facilitate conversations between research and IT staff within their Area when considering regulated data environments.

**Summary**

WSU faculty have expressed concerns and raised questions with the proposal for a centrally maintained regulated data environment (RDE). This document is intended to address that feedback and describe scenarios that have been experienced in the past by research faculty. Additionally, it provides a functional outline for building a solution that meets WSU’s research enterprise and other needs in situations where regulated data is required.

For a research project involving regulated data, the terms of the contract, data sharing agreement, business associate agreement (BAA), or pre-IRB assessment dictate the environment in which that data must be accessed, processed, and by what means that data can be shared. The scale and complexity of meeting such requirements range and differ based on which named framework is required.

Meeting the documentation, security, and compliance requirements of a regulated data framework is often a significant obstacle. A centralized model offers an opportunity to more efficiently provide solutions that meet named regulatory frameworks relative to independently configuring such infrastructures within individual campuses, areas, or colleges. The time, cost, and other resources needed to be fully compliant with a particular regulatory framework are significantly reduced with centralization while simplifying the ability to secure and monitor hosted environments.

Such a properly architected system will allow for access and availability to a regulated environment from anywhere in the world for a multitude of data processing tasks; such as genomics, artificial intelligence (AI), machine learning, and other high-performance computing workloads. Whether you are in your office or working remotely, the centralized data environment will allow you to access your data safely, regardless of size of the task.

Accessibility for cross-collaborators external to the University, the process by which they are initially provided access, and how they interact with a WSU maintained regulatory data infrastructure is under consideration.

**Current State**

Regulatory data environments currently exist independently across the institution, within a handful of campuses, areas, and colleges, in which contractual obligations have necessitated their implementation. In such instances, each unit has borne the cost and resources required to facilitate a compliant environment.

In addition to unit specific environments, WSU’s only centralized regulatory data environment, at this time, is a NIST SP 800-171 compliant tenant in Amazon Web Services. This environment provides a solution for management and processing of data that requires the use of that framework. Outside of this implementation, there is no current centralized model for provisioning other regulated data environment quickly and efficiently.

**General Functional Needs**

WSU researchers often encounter regulated data when working with clinical, community-based, qualitative, and defense-related research.

Each of those research scenarios has specific requirements related to how regulated data is processed and stored. There are also use cases where there is a blending of the scenario examples as well.

When performing research that involves regulated data, RDE requirements that most likely must be addressed in all scenarios are noted as follows:

* Scope
	+ Prescreening prior to award or approval
	+ eConsent
	+ Support for collaborators/affiliates
	+ Statistical software and productivity tools
	+ Secure survey tools and offline support
	+ Access from anywhere
* Cost/Scalability
	+ Scalable
	+ Highly available
	+ Active and cold storage
	+ Possible HPC needs
	+ Custom solutions/development - Ability to for WSU software developers to create solutions using a variety of programming, analysis, and database tools
* Security
	+ Auditability
	+ Data lifecycle management
	+ Role based access
	+ Ingress and egress controls
	+ Secure workstation/s
	+ Secure data transfer (such as SFTP or Globus)
	+ De-identification process that allows researchers to get confidential/regulated data into a format that is approved for analysis externally
* Support
	+ End-user support
	+ Infrastructure support
	+ Cost model/billing

**Scenario - Clinical**

In this scenario, researchers are studying human subjects and collecting information pertaining to them. Because of the data collected, there is a need to provide e-consent, formalized survey/collection tools, securely store the data collected for the life of the project and analyze the data in a secure environment. Access to this environment is not bound by location. This research can vary in size from a single PI, an entire lab, to cross functional labs spanning multiple disciplines. As permitted, the ability to export results is desired. Data lifecycle management is desired.

* Unique focus on data collection/survey tools
* Connection between REDCap and secure space will be a must for many clinical research projects.
* Researchers that are collecting human subjects data in the format of recorded interviews. Data needs to be captured securely, and then have an environment to transcribe and remove identifiers (not all projects will remove them) so qualitative analysis can be done elsewhere.

**Scenario – Community-based**

In this scenario, researchers spanning multiple colleges, campuses, or institutions (collaborators) are working with data provided by a 3rd party or data that is self-generated via surveys. Being able to securely obtain data from a 3rd party while adhering to the requirements outline by the data sharing agreement. There is a need to allow access to WSU resources by non-WSU collaborators. Because of the data used/collected, there is a need to provide e-consent, formalized survey/collection tools, securely store the data collected for the life of the project and analyze the data in a secure environment. Access to this environment is not bound by location. As permitted, the ability to export results is desired. Data lifecycle management is desired.

* Unique focus on collaborators/affiliates. There is a need for a straightforward solution to add affiliates so there is a formal method for access using WSU assigned credentials.
* Additional needs for collaboration tools to address data sharing, messaging, email, etc.

**Scenario - Qualitative**

In this scenario, researchers are studying data with minimal (if any) exposure to the subjects. There is a need to store 3rd party data while adhering to the requirements outlined by the data sharing agreement. Because of the data used, there is a need to securely store the data collected for the life of the project and analyze the data in a secure environment. Access to this environment is not bound by location. This research can vary in size from a single PI, an entire lab, to cross functional labs spanning multiple disciplines. As permitted, the ability to export results is desired. Data lifecycle management is desired.

* Unique focus on 3rd party data and meeting the requirements of a DSA

**Scenario - Defense-Related Research Involving CUI**

In this scenario, researchers are working on a subcontract with requirements related to Controlled Unclassified Information (CUI). Due to the potential application of the research to real-world military designs, the federal agency designated nearly all government material provided to the contractor and subcontractors, as well as the results generated by the contractor and subcontractors, as CUI. There is a need for university researchers to securely access the information provided by the government, as well as to generate and analyze research results in a secure environment that is compliant with all CUI regulations.

* Additional needs surround licensing requirements of specialized software. If cloud computing is possible (assuming sufficient funding is built into the subcontract), the question remains how software licenses would transition to the cloud environment.
* With Cybersecurity Maturity Model Certification (CMMC) requirements being phased in for defense research, secure environments (or the plan to at least readily implement one) may need to exist at the solicitation stage.

**Additional comments regarding regulated data scenarios**

Secure collaboration space. Researchers ask how to accomplish this, especially to serve a team spanning multiple campuses and disciplines. This use case could be needed for clinical, community, and qualitative scenarios.

We will still have projects that may or may not work within a cloud-based environment (e.g. where specific DUAs do not permit the use of networked computers), this can be common for projects utilizing DOJ/NIJ, some Department of Education data etc. These will need to be handled separately on a case-by-case basis.

A major IRB/ATO challenge is assessing the security of the constant stream of new programs faculty want to use to do remote/on-line research. Not all of the programs faculty will want to use will work within, or in connection with the secure environment.

Setting up eConsent that complies with FDA/HHS/NIH regs alone is not that hard to work out, what our faculty are struggling with is having limited access participant portals that support eConsent with MFA validated signatures (required to be considered legally valid in WA).

**Recommendations**

* Scope
	+ Identify the regulatory data environments available as part of this platform based on existing and anticipated future utilization (eg: HIPAA, NIST 800-171, NIST 800-172, WA OCIO 141.10, GDPR, CMMC, FISMA, FERPA, etc.)
		- Upcoming Washington State legislative bills regulating privacy rights (eg: CCPA type regulations) should be closely monitored and anticipated.
	+ Provide template environments for meeting specific regulatory data frameworks.
	+ Identify the processes and requirements for ensuring that external cross-collaborators can access the regulated environment solution.
		- Current process for external access uses an “Affiliate” account created within WSU infrastructure. Alternative and simplified processes should be considered and developed if possible (eg: additional Single Sign-On integrations, support for multiple MFA solutions).
	+ For proposals that may include regulated data, pre-assessments should be made to determine the costs associated with operating within the tenant.
	+ Identify and provide communication platforms that can meet supported regulatory data frameworks (eg: email, voice, video, text).
	+ Consider adjustments to the IRB process to better address regulatory data requirements during pre-assessment.
	+ Consider options for automation of data migration to cheaper storage environments when not in use or for retention purposes.
* Cost/Scalability
	+ Determine and provide base compute and storage costs.
	+ Provide a metric for identifying costs based on anticipated data processing and management; particularly for HPC workloads.
	+ Provide recommendations for utilization of compute availability process types and costs associated with those options (eg: dedicated, shared, spot).
	+ Develop a model and process for billing distribution and invoicing for tenant users.
* Security
	+ Identify privileged and identity access management processes.
* Support
	+ Staff a centralized representative at ITS to help researchers facilitate the planning, implementation, and onboarding process within the centralized regulated data environment.
	+ Create reference documentation to familiarize faculty and IT support units with the available options.
		- Tenant organizational diagram.
		- An example of a bill and how charges are itemized.
		- Guidance for when a tenant experiences unexpected compute costs and how to address them.
		- Best practices guidance.
	+ Create reference documentation for how regulated data environments will be accessed, how data will be moved in, moved out, and shared among collaborators.

**FAQ**

* How does our existing Office 365 infrastructure tie into a centralized regulatory data environment model and what Office 365 services (eg: Office suite, SPO, Teams, OneDrive, approved add-on apps) does that cover if included?

*The WSU Office 365 environment is not currently within the scope of the proposed general RDE solutions. There are efforts underway to certify the Office 365 for the processing of ePHI data which when completed and associated controls adopted by staff/faculty will allow certain select regulatory requirements to be met.*

* Who will assist research faculty with planning and implementation processes and, if a dedicated support person or team, how will that position be funded?

*ITS has committed existing resources to assist with the planning and implementation processes but the ongoing operational support and associated positions have not been funded.*

* If researchers are coordinating through a support person or team for integration within the centralized environment, will there be costs associated with such support, and if so, what is the anticipated rate?

*There will be no ITS labor charges for the initial consultation and design/deployment of an RDE.*

* Where does the burden of support, maintenance, and monitoring lie?

*Ongoing operational support and maintenance, including monitoring will lie with the research unit.*

* How will data retention requirements be addressed upon completion of a research project?

*In accordance with the data management plan.*

* What options will be available for backup and recovery within regulated environments? What are those costs?

*Costing services in an Infrastructure-as-a-Service (IaaS) environment are complex and prone to error. Faculty members should attend WSU Cloud Computing Office Hours to build a custom cost calculator around their needs.*

* What capabilities/options/limitations are there for integrating on-premise equipment and instrumentation utilized for research tasks?

*Integration of existing, on-premise infrastructure is not within scope of the RDE. Any costs associated with implementing and meeting regulatory data requirements to accommodate on-premise equipment will lie with the research unit.*

* What options are available for automated deidentification of datasets and review?

*Deidentification is outside the scope of the RDE.*

* How will RDEs address historical slowdowns when having to utilize a centralized data environment?

*ITS is utilizing proven cloud service providers to implement the underlying infrastructure for the RDEs. As such, a researcher can provision as much, or as little compute, RAM, and speed of storage as they can afford. The network latency for the five main campuses to the cloud are sufficient for the virtual desktop interface (VDI) that researchers will use to access the RDE and the consultation engagement phase will determine whether bandwidth requirements can be met.*

* How will the RDEs support remote field work?

*As long as the remote field worker has access to Wi-Fi, cellular, or other broadband network options they will be able to access the RDE.*

* How will the RDEs support colleagues and collaborators outside the WSU system?

*The RDEs will support the creation of userid/password credentials different from the WSU Network ID (NID) and will support the creation of credentials for non-WSU personnel as well.*