Registries Project Proposal Framework

DRAFT

This document is still a work in progress.

Prepared by:

Document Information and Revision History

1 Introduction

1.1 Background

Participants from a number of organizations have been collaborating on creating a coherent set of open-source Identity and Access Management (IAM) software packages to meet the needs of Higher Education and Research. In the development phase this initiative is being called "Open Source Identity Management for Higher Education" (OSIdM4HE). The activity arose in response to concerns raised by many institutions that current products, both open-source and commercial, are not meeting their IAM needs effectively and affordably. Several weeks of analysis have identified requirements in a number of functional areas, and an assessment of some candidate open-source products. The group has also developed proposals for organizational and funding structures to support development and integration projects. Below are some important concepts and assumptions to keep in mind while reviewing this document

1.2 Purpose of this Document

The purpose of this document is to outline the necessary pieces and functions that the Registry portion of the overall solution would require. In addition, it outlines the visions, objectives, and how the project could hypothetically be structured if the Kuali Foundation were set as the "CareTaker" organization.

1.3 Key Terms and Definitions

Term	Definition

Affiliation	Affiliation is the combination of one's relationship with an institution (which may allow access to electronic services) and some form of trusted (may not be University) identity.
Assurance Levels	Assurance levels are numerical levels (or degree levels) that correspond to the degree of confidence in the vetting and proofing processes used to establish the identity of the individual to whom the credential is issued (Levels of Assurance)
Attribute	Information associated with a digital identity record. Attributes may be general or personal. A subset of all attributes defines a unique individual. Examples include name, phone number, and group affiliation.
Levels of Assurance	The degree of confidence in the vetting and proofing processes used to establish the identity of the individual to whom the credential is issued. Levels of Assurance also consider the degree of confidence that the individual who uses the credential is the individual to whom the credential was issued.
Vetting	The process of validating information, collected from an individual, for the purpose of issuing digital credentials.
Credentialing	The act of issuing a token that will be used to establish a digital identity.
Proofing	The act of aligning a person's previously recorded data to the actual person at the time when credentials are issued. In-person proofing involves checking a photo ID, such as a driver's license, against the holder of the ID.
Affiliate	A person who has some connection to the University.
Net Id	An electronic identifier created specifically for use with on-line applications.
Identity Assurance Profile	A set of data, associated with an individual, that reflects the degree of confidence in the vetting and proofing processes used to establish the identity of the individual to whom the credential is issued at a given point in time. See "Assurance Levels" and "Levels of Assurance" for related information.

2 Project Objectives

2.1 Vision

The function of an identity registry is to register and maintain information about entities of interest to the organization operating the registry, and to make this information available to other systems. To this end, it plays an important role within the enterprise where it can serve as the authoritative source for identity information.

Currently, within the open source space, there are not many options for an identity registry solution. Many have chosen to leverage Oracle's solution (formerly Sun) *include more details here* while others have decided to assemble their own. There is also the Open Registry project from JA-SIG which was started at Rutgers.

Within Kuali, the Kuali Identity Management (KIM) module of Kuali Rice provides many features of an identity registry, but lacks some critical pieces of registry functionality. The introduction of the Kuali People Management for the Enterprise (KPME) project (which implements a comprehensive HR/payroll system) as well as continued work on the Kuali Student system has continued to drive requirements that align with the identity registry space within Kuali. As a result of this, work that is being done within Kuali has the potential to result in an identity registry solution that could be used outside of the context of the other components of technology being built within the Kuali Foundation. This would potentially be of benefit to any institutions who chose to adopt such a solution either as a standalone solution, as part of the adoption of other Kuali solutions, or as part of the proposed OSIDM4HE "suite".

With these thoughts in mind, a vision for an identity registry solution can be set forth:

⚠ **TODO**

Fill in details on vision, need to reconcile it with the objectives listed below as there will likely be some overlap

2.2 Project Objectives

The specific objectives of each group participating in the development of the Registry may be very different. However, the common objectives that will achieve the vision of the project can be stated as follows:

- To deliver a registry product that could be use as part of the overall OSIdM4HE suite of products or on it's own in conjunction with existing products or services a higher education institution may employee.
- The registry and its functions will also be built in a modular fashion to allow institutions to utilize only the pieces it may need.
- The registry will be built
- The registry will be developed and built in such a fashion that it may be implemented in an "above the campus" or cloud structure.
- The application will be community sourced where investing partners have the initial say in the direction of the project, but contributions and input are welcome from the open source community as a whole.

2.3 Key Success Indicators

The following key success indicators have been defined in order to measure the success of the Kuali Rice Project. They are not listed in any relevant order of importance

1. The cost of executing the Registry project are within the expectations of the Investing Partners.
2. Case studies and conference presentations of successful implementations are regularly appearing on mainstream conference programs such as Campus Technology, EDUCAUSE, NACUBO, and ACCRAO.
3. All Registry Investing Partners have implemented core components of the Registry

4. Adoption meets and exceeds targets set by the Board for each major release.
5. Healthy community support activities and resources are available for adopters. The Registry community is developing and sharing materials and documentation. Community support activity on mailing lists and other forums is timely, vibrant and healthy.
6. A healthy vendor and affiliate community exists around the Registry that includes consulting companies and support organizations. Institutions are able to find implementation assistance that is timely and effective. The Registry and/or its components are integrated into popular vended and open source software infrastructure stacks and institutions are able to secure commercial support that complements information from the community.
7. Re-usability of components, and flexibility to adapt to business processes from multiple organizations.
8. IP and license compliance. All contributor license agreements (CLAs) have been signed before development starts and there is 100% compliance to the CLA's and 3rd party licenses before each software release.
9. Registry passes the “not invented here” (NIH) test. The investment in the Registry is consistently used for the highest value items unique to our education community. Re-use of world-class open source software components from communities outside of education, and participation in these communities is a core competency of our community that increases the breadth and depth of what the Registry offers, increases the quality and quantity of talent, and results in software that proves high-quality, scalable, and sustainable.
10. Vendors and institutions have been encouraged and empowered to test and certify the Registry to run in various combinations with different technology stacks in addition to the provided reference implementation.

3 Functional Scope

3.1 The Registry Functional Model

This model is concerned with identity registries serving institutional needs: containing thousands or millions of entities, operated according to institutional policies to meet institutional goals such as accountability, compliance, security, and collaboration. Requirements for identity registry functionality and operation are derived based on the terms and concepts presented in this model.

Entities, entries, identity, identifiers - An entity is a "thing" of interest to the institution, distinguishable from other entities of its type. Entities of most interest for an identity registry are typically "actors", i.e. things that initiate actions in online systems. The most common type of entity is a person, hence identity registries are often called person registries. Other common "actor" entities are processes, applications, computers, and organizations. An entity is represented in the identity registry by a record called an entry that contains structured information about the entity. Some of the data describes the entity; this is identity data. Other data, such as entry create time or access control data, is registry metadata. A data element that is designed to distinguish entities in a set is called an identifier. An entry typically contains several kinds of identifiers, as well as other data about the entity. A key goal of a registry, typically, is to ensure, as much as possible, that each entity is represented by exactly one registry entry. Each entry in a registry has a type, and each type has a schema. Different types may be handled by different registries, or a single registry may deal with several types.

Registry-managed identifiers - In addition to managing entity data sourced from various business processes, identity registries typically are source systems (i.e., are authoritative) for some data, in particular institutional identifiers. A common registry-managed identifier is a registry ID (also called unique ID, or UUID) that is an opaque non-reusable identifier serving as an institutional "key" for the entity. Another common registry-managed identifier is a network ID (also called NetID or username) that is used by end-users for login and other services such as email. Creation and management of NetIDs, and other similar identifiers such as Distinguished Names, may be integrated with credential assignment processes that also include management of authentication information such as passwords and public keys. This is a touchpoint between identity registries and authentication systems.

Registration, matching, reconciliation - Registration (also known as enrollment) is the process of creating a new identity registry entry. Identity data may come into a registry from source systems (which are typically also registries in a sense), or interactively via human entry processes. A person who engages in registering entries is called a registration agent. In support of the goal of one entry per entity, it is necessary for the registration process to determine whether a set of identity data coming into the registry refers to an existing entry, or represents a new entity, hence requiring the creation of a new entry. The process of distinguishing new from existing is called matching. The matching process may rely on many different data elements, and may involve human decision-making in addition to automated processing. The process of adding or modifying identity data in an entry based on incoming data is called reconciliation.

Merging, splitting - It may be found that due to a failure of matching in the registration process more than one registry entity exists for an entity. In this case two or more entries must be merged. Similarly, it may be found that an entry contains a mix of information from different entities. In this case the entry must be split into two or more entries. Merging and splitting are typically administrative processes; in the case of person entries the processes may involve active participation of the affected people.

Identity information distribution - Information in identity registries is made widely available to many processes and systems to support institutional-scale identity integration. This implies a touchpoint between identity registries and information distribution (or presentation) systems such as directories, web services, etc. This may imply requirements for low latency of change propagation, high fanout, etc.

Affiliations, lifecycle - Many different institutional processes bring entity information into a registry. In addition to the entity's type (person, e.g.), the registration process and the information in the entry typically reflect the nature of the process that brought the entry in. For example, the entry for a person who is a student will likely have a different input process and hold different information from that of a person who is an employee (a person may be both, of course). The different relationships that affect entry data and maintenance are called affiliations. The policies and procedures that codify how an entry is managed over time are called lifecycles of the various affiliations. Identity registries may need to support affiliation catalogs, representations of lifecycles and policies, and integration with business systems such as student and human resources systems that implement the lifecycles for key institutional affiliations. Affiliation lifecycles also often affect service access by people and other actors, implying a touchpoint of affiliation and lifecycle management with access management systems.

Contact / profile information - A common class of identity data is contact information, or more generally profile information. Contact information includes items such as phone numbers, email addresses, web URLs, etc. Profile information may include many types of information including departmental associations, interests, and other relatively unstructured information.

Identity assurance - Institutions have a wide range of relationships with people, from rich and long-term to brief and casual. Information about people in an identity registry may be well-managed and vetted by trusted business processes, or it may be self-asserted or untrustworthy, depending on circumstances. This consideration leads to notions of representing these qualities of information, a concept called identity assurance (registry information is only one piece of overall identity assurance). Identity registries may have requirements for storing and managing assurance-related information, such as vetting processes, times/locations of data checking, etc. This also implies touchpoints with other aspects of identity assurance, in particular authentication systems.

Management operations / user access - Identity registries are maintained via many processes; some of these involve interactive access by registry-associated staff to do such operations as status checking, handling merges and splits, investigating data anomalies, viewing entry history, generating reports, etc. This implies requirements for user interfaces to support these operations, including appropriate access control. Other functions may imply a need for end-user access for operations such as profile management, self-service merging, service requests, etc.

3.2 Functional Scope Overview

> the sub-sections below currently contains a dump of the sub-group's requirements, it should be summarized into 7 or 8 summary requirements for easier digestion

3.2.1 Registry

The registry is the core data store of identity information. What follows are requirements related to it.

Requirement ID	Requirement Description
REG_0100	The registry shall support the storage of identity information.
REG_0110	The registry shall support the storage of the partial (MM/DD) and/or full (MM/DD/YYYY) date of birth for a person.
REG_0120	The registry shall have a unique identifier (non-SSN) for each person in its data store.
REG_0130	The registry shall support the storage of a person's gender.
REG_0140	The registry shall have the ability to storage multiple net ids for a person.
REG_0150	The registry shall have an indicator as to which is the primary net id for a person.
REG_0160	The registry shall store a person's first, middle (optional), last name and suffix (optional).
REG_0170	The registry shall maintain a history of a person's name changes.
REG_0180	The registry shall store a type associated with each person's name (for example, legal name).
REG_0185	The registry shall have the capability to store multiple name types for a person. Examples include: legal name, preferred name, ...
REG_0190	The registry shall support the storage of multiple addresses for a person, indicated by a type (for example, employee home address).
REG_0200	The registry shall store for an address, the following information: street address (multiple), city, state, postal code, country, campus location and source.
REG_0210	The registry shall store for a person's name an flag to indicator whether a first name is unknown (FNU) and/or a last name is unknown (LNU).
REG_0220	The registry shall support the storage of multiple telephone numbers, indicated by a type (for example employee office telephone number).
REG_0230	The registry shall store for a telephone number the following information: area/country code, phone number, extension (optional) and source.
REG_0240	The registry shall support the storage of a person's email address(s) and their respective type.
REG_0250	The registry shall maintain a history of all of a person's address changes.
REG_0260	The registry shall maintain a history of all of a person's telephone number changes.
REG_0270	The registry shall maintain a history of all of a person's email address changes.
REG_0280	The registry shall support the storage of information about all of the credentials a person holds (for example: kerberos principal, secure id token serial number, PKI, ...).
REG_0290	The registry shall support the storage of all a person's affiliations.
REG_0300	The registry shall support the storage of a person's Identity Assurance Profiles.
REG_0310	The registry shall store information related to an identity proofing event.
REG_0320	The registry shall support the storage of identity card information.
REG_0330	The registry shall either store an indicator or have a calculation to determine a person's primary affiliation.
REG_0340	The registry shall support the mapping of its affiliations to the eduPerson attributes.
REG_0350	The registry shall provide a comments facility to be used for authorized personnel (Security) to record information about person's identity.
REG_0360	The registry shall have complete auditing of information in its registry
REG_0370	The registry shall provide a facility by which authorized personnel can obtain a read-only view of portions of its data.

REG_0380	The registry shall maintain a single namespace for person identifiers and network ids.
REG_0390	The registry shall support the storage of common HR and student information, like title, status and department.
REG_0400	The registry shall support the storage of international forms of user information.
REG_0410	The registry shall support the storage of organization-specific attributes.
REG_0420	The Registry shall support the ability to associate a START DATE with each Affiliation type.
REG_0430	The Registry shall support the ability to define different life cycle processes for removing an Affiliation type (eg staff accounts are disabled immediately; faculty retain their accounts for six months but with reduced services).
REG_0440	The registry shall support the ability to associate an END DATE with each Affiliation Type.

3.2.2 Identity Merge

Despite the best of tools to identify and prevent the entry of duplicate identities into an Identity Registry, duplicate identities will inevitably be created. The Identity Merge facility will provide the ability to link or merge identity data to a single unique identifier.

Requirement ID	Requirement Description
MERG_0001	The Identity Merge facility shall provide the ability to link unique identity records within the registry identifying one as the primary.
MERG_0002	The Identity Merge facility shall maintain for historical purposes the non-primary registry entries. These entries shall not be available for further use (other than historical reporting) in subscribing systems.
MERG_0003	The Identity Merge facility shall provide the ability to merge identity data to a "primary" registry entry through either an automated process or a user controlled process (eg. system identifies records to merge, shows user relevant data, user decides interactively what to accept etc.)
MERG_0004	The Identity Merge facility shall provide notifications to subscribing systems of the merged identities
MERG_0005	The Identity Merge facility shall provide for the creation of business rules for the automated merging of registry data elements. (eg. defining the authoritative source by attribute when duplicates are identified)
MERG_0006	The Identity Merge facility shall be integrated with the Access Management Module to allow for assessing and potentially merging role and access data when duplicate identities are identified and merged.
MERG_0007	The Identity Merge facility shall be integrated with the Provisioning Module to allow for the potential deactivation of duplicate accounts.
MERG_0008	The Identity Merge facility shall use a fuzzy logic searching capability for matching purposes.

3.2.3 Management Functions

The section will detail requirements related to management information that the registry should provide. The interfaces will typically be Web Services (SOAP and/or REST-based).

Requirement ID	Requirement Description
MAN_0100	The registry shall provide interfaces for authorized registry authorities to manage information in its data store.
MAN_0110	The registry shall provide services to add, update, and archive persons.
MAN_0120	The registry shall provide services to add, update, and archive address information for a person.
MAN_0130	The registry shall provide services to add, update, and archive name information for a person.
MAN_0140	The registry shall provide services to add, update, and archive telephone number information for a person.
MAN_0150	The registry shall provide services to add, update, and archive net id information for a person.
MAN_0160	The registry shall provide services to add, update, and archive credential information for a person.
MAN_0170	The registry shall provide services to add, update, and archive Identity Assurance information for a person.
MAN_0180	The registry shall provide services to add, update, and archive affiliation information for a person.
MAN_0190	The registry shall provide services that are either SOAP and/or REST-based.
MAN_0200	The registry shall provide a web-based front-end to the data contained in its registry for authorized personnel.
MAN_0210	The registry shall provide a flexible batch-file interface for importing and extraction of data, including support for fixed column, CSV, XML, .xls, and other formats.
MAN_0220	The Registry shall provide a means of purging categories of entries (eg Applicants)
MAN_0230	The registry shall provide a configurable means by which changes can be (optionally) reviewed and approved

3.2.4 Enterprise System

This section will detail requirements related to the enterprise system aspect of the registry.

Requirement ID	Requirement Description
ES_0100	The registry shall support the notification of data changes to entities either using publish/subscribe or point to point communications.
ES_0110	The registry shall support auditing of all actions for a person record.
ES_0120	The registry shall support data reporting of registry data for authorized personnel.
ES_0130	The registry shall notify end-users via email *x* days prior to the expiration of their services.
ES_0140	The registry shall have rules for cleansing and standardizing data before its entered into the data repository.

3.3 Scope Determination and Decision Making Process

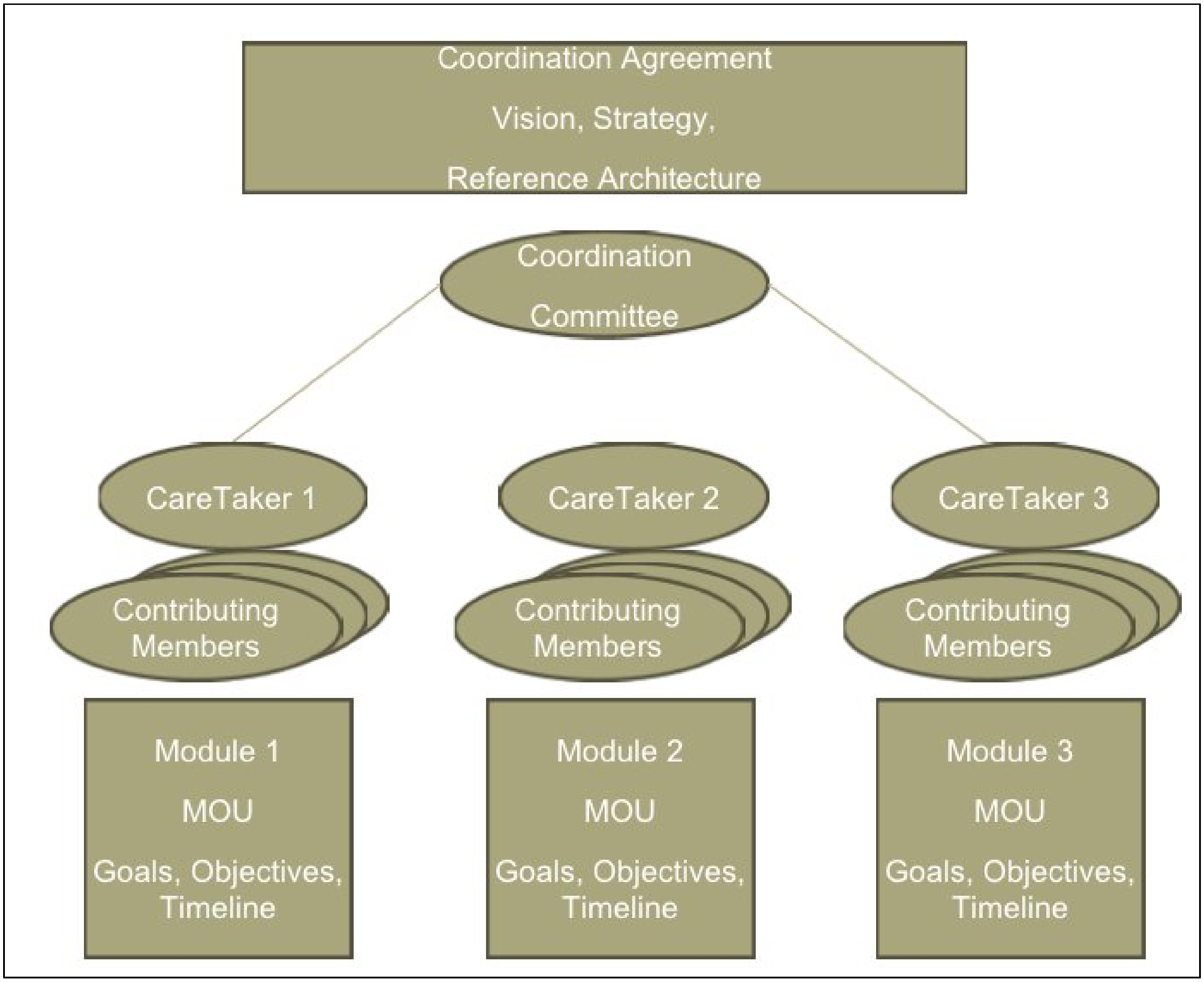
4 Technical Architecture

> this section should contain the discussions we've had as a group regarding reference architecture, cloud based frameworks, etc.

4.1 Overview

4.2 Architecture Direction and Decision Making Process

5 Project Organization



5.1 OSIAM4HE Coordination Committee

Responsibilities:

- Act as the coordinating body responsible for overseeing the production of a new IAM suite of open source software.
- Creating and protecting a “Coordination Agreement”, to be read and signed by each contributing member, that describes the overall OSIAM4HE vision, strategy and organization structure
- Documenting a reference architecture, and interoperability standards, that is to be adhered to by each module to ensure continuity between modules
- Meeting on a recurring basis to review progress, deliverables and timelines of each module and to help resolve potential conflicts
- Producing communications to the general public as to the goals of the effort, means of participation and progress toward module completion.

5.2 Module Team

- The Registry module will represent a discrete set of IAM functionality that has been identified, agreed upon and prioritized by the OSIAM4HE Coordination Committee
- The Registry module will have a MOU following a common format that identifies among other things:
  - Which organization is the “Primary Caretaker” with any specific requirements that the caretaker requests of contributing members
  - The individual that will serve as the designated liaison on the OSIAM4HE Coordination Committee
  - High level goals, deliverables and timelines of the module
  - Expectations of contributing members including the resources, roles and responsibilities being contributed
  - The standard module administrative practices and development procedures to be followed

5.3 Primary Caretakers

Kuali, specificially the Kuali Rice project, will be the primary caretaker for the Registry module. Responsibilities:

- Primary Caretakers of each module shall have responsibilities for long term sustainment of the modules code base
- Primary Caretakers may seek and use different contributing members (and/or affiliates) than those involved in the initial MOU development for the long term sustainment
- It is assumed the Primary Caretakers are an established non-profit entity with higher education community support, legal and financial mechanisms in place that protect the long term viability of the module development

5.4 Partnerships

5.4.1 Investing Partners

TBD - What are the investment levels?

5.4.2 Other Partners

TBD - Are there other types of partners? What would they look like?

6 Project Team Organization

With Kuali Rice as the module caretaker, the project team will consist of the following roles:

- **Kuali Rice Project Manager**
- **Kuali Rice Lead Architect**
- **Kuali Rice Business Analyst**
- **Kuali Rice UX Architect**
- **Registry Team Development Manager**
- **Registry Team Developer(s)**

7 Project Management

7.1 Project Planning

Project plans are critical to providing the required framework from which to control the project and ultimately facilitate its successful outcome. It is critical that the plans are well understood by all project team members and stakeholders. Before each new development phase begins, the work plans will be reviewed and adjusted based on the knowledge gained from previous development phases.

7.2 Project Communications

Status reporting and communication is an essential component of project management and control. It is an effective mechanism for communications *within* the project and to associated projects. The table below outlines the proposed status reporting schedule and communication strategies.

7.3.1 Project Status Report

The Project Manager will prepare a bi-weekly status report for the project that will inform the following groups about project status:

- OSIAM4HE Coordination Committee
- Module Team
- Kuali Rice Board

This status report will include:

- Percent completed on planned activities
- Planned / Actual/ Variance reporting on project costs
- Upcoming project activities
- Highlights of any project issues or escalation points
- Risk Report and Forecast

7.3 Change Request and Issue Management

TBD

7.4 Risk Management

TBD

7.5 Project Documentation Strategy

Specific documentation will be prepared for each of the project written deliverables.

7.6 Collaboration Tools

Effective communication and collaboration is a critical success factor in any project, but it is particularly important within a distributed project with remote development nodes. While the project will try to minimize travel requirements due to cost and personal impacts, some travel will be required to ensure the success of the project and to ensure that communication paths remain open. It is essential that the Registry Project have effective tools to support collaboration efforts. The following Kuali tools will be used to support remote collaboration efforts:

- Video and audio conferencing through bridge technologies
- Instant messaging services (such as Jabber and Skype)
- Adobe Connect – net meeting product
- Confluence – Wiki for sharing materials
- JIRA – bug tracking and issue tracking
- Google Apps – mailing lists, calendar, document sharing