pennsylvania
OFFICE OF ADMINISTRATION
INFORMATION TECHNOLOGY

Information Technology Policy

Directory Services Implementation Guide

ITP Number	***Effective Date***
BPD-SEC013H	September 7, 2006
Category	***Supersedes***
Recommended Policy	
Contact	***Scheduled Review***
RA-ITCentral@pa.gov	Annual

1. Introduction

The purpose of this documentation is to provide best practices for implementation of the architectural requirements and services defined in supporting document GEN-SEC013B - *Directory Services Architecture*, concerning a Commonwealth of Pennsylvania Enterprise Directory (CoPED).

2. Distributed Model and Technology Overview

As SEC013B explains, CoPED consists of a compilation of identity data supplied by the various agencies. The architectural model selected for CoPED is a Distributed Model comprised of three logically separate domains: *Employee*, *Business Partner*, and *Subscriber*. The identities stored in CoPED are to be users of commonwealth services and resources, not businesses or artifacts of any other type, only persons. These identities are not restricted to a single domain, but may exist as a separate account, with separate ID numbers and credentials, in each domain. No person is to have more than one identity in each domain, and agencies are directed to apply due diligence to prevent such occurrences from happening.

There is no single best implementation strategy for CoPED. As is the case with most large endeavors, CoPED will be rolled out incrementally. Each agency is to adopt the strategy that makes the most sense for its specific needs and available resources. Several tools have been established to support these implementations:

- Virtual directory technology has been selected for agencies to create integrated “views” of the three domains – composite data sets of entities and attributes to support their own agency-unique application and service requirements. Integrated security mechanisms prevent unauthorized users from viewing or editing restricted entities or entity attributes as prescribed by the data’s authoritative source or governing agency.
- Directory technology has been selected not only for implementation of the Commonwealth Enterprise Directory, but also for agencies to establish their own Master Directories. Master Directories will allow agencies to consolidate their own internal data stores, and simplify synchronization of their identity data with CoPED.
- Metadirectory technology has been selected to provide data transformation and synchronization functionality that will empower agencies to create and

manage new master directories. Metadirectory provides the engine for synchronizing these master directories with CoPED, while preserving the physical separation of the three CoPED domains. It can be leveraged to create and synchronize agency directories that need to be disconnected from their networks, or to combine locally stored identity attributes with CoPED data to create unique directories for specialized application requirements. Section 8, *Virtual Directory Guidance*, addresses synchronization in detail.

3. CoPED Implementation Strategy

3.1 Employee

Comprehensive employee directories already exist within the enterprise. CoPED will leverage this existing data by presenting a virtual view of identities and their attributes in one of these approved identity stores, rather than storing an additional copy of the data in a new instance. CoPED will use the Commonwealth of Pennsylvania Active Directory forest (CWOPA) as its initial *Employee* directory, providing a nearly complete listing of employees under the governor's jurisdiction. Although CWOPA also contains non-employee records (mostly contractors), virtual directory technology can be used to hide or filter out these accounts.

One open issue is the fact that there exist employees who do not have accounts in CWOPA. Initially, CoPED can simply ignore these entities, but they will eventually need to be included for a complete *Employee* domain. There are two logical methods to accomplish this: the first and probably easiest (though possibly most expensive) method would be to systematically add these accounts to CWOPA. While this may be the simplest solution to implement, it would probably also be the most expensive because the commonwealth would have to buy Windows Client Access Licenses for each of them. The second method would be to create a new *Employee* directory instance, similar to the *Business Partner* and *Subscriber* directories, and utilize metadirectory technology (identity synchronization) to populate with data from the SAP HR system (IES) which contains the definitive list of all Commonwealth employees. The correct option is to be made according to the criticality of the business requirement for having these "missing" employees in CoPED.

3.2 Subscriber

The *Subscriber* domain consists of those persons who are enrolled to receive commonwealth services. Primarily composed of state residents, this domain will likely also include some people who are out-of-state. While there are many existing subscriber identity stores throughout the commonwealth, a new directory will be built to support the goal of non-duplicity (one account per user) and the attributes required for CoPED.

Several agencies have comprehensive citizen identity stores that may be leveraged as data sources for the initial transfer of citizen data to CoPED. This transfer would be subject to the identification of each user by a globally unique identifier (GUID). These entities will be flagged to indicate the source of the

record, and to note that they have not yet been vetted under the current commonwealth requirements. This flag will change when the citizen is properly vetted as described in the supporting document, GEN-SEC013G - *Public Key Infrastructure*.

The commonwealth's existing Web presence includes an Active Directory forest for self-registered users, called *User*. This architecture will not include that directory in its initial implementation, and it is not expected that those entities will be migrated by any CoPED or other IPAM process into the *Subscriber* directory.

3.3 Business Partner

The *Business Partner* domain includes all accounts that belong in neither the *Employee* nor the *Subscriber* domain, and consist of what are typically described as non-employees who provide services to or on behalf of the commonwealth. These identities may have other accounts in the *Employee* or *Subscriber* domain if they are employees or subscribers of commonwealth services. The *Business Partner* accounts will initially include non-employee first responders who need commonwealth-issued PIV cards for interoperable federal identification, and contractors who are in CWOPA. Later additions are expected to include county and municipal workers as well as other types of contractors and vendors.

As noted earlier, one goal of IPAM is for each entity to have no more than one account in a given domain, even if they have accounts in the other domains. Within the *Business Partner* domain, significant hurdles exist due to the varying nature of how these identities are created and used by the different agencies:

- The same business partner might be used for different purposes.

- Organizations contain multiple individuals, and while each of those individuals might exist as separate entities, agencies might also be identifying the organizations themselves as entities (IPAM is intended to only contain user identities, not organization identity).

- A single individual might have multiple roles that make it difficult to record a single entity for that individual.

- A single subcontractor may be a member of multiple teams or multiple businesses.

Therefore, although the goal is for each individual to exist as a unique entity in the *Business Partner* directory, duplicate identities are permissible when necessary to address pragmatic needs.

The commonwealth's existing security infrastructure includes an Active Directory forest for managed non-employee users, called *MUser*. This

architecture will not include that directory in its initial implementation, and it is not expected that those users will be migrated by any CoPED or other IPAM process into the *Business Partner* directory.

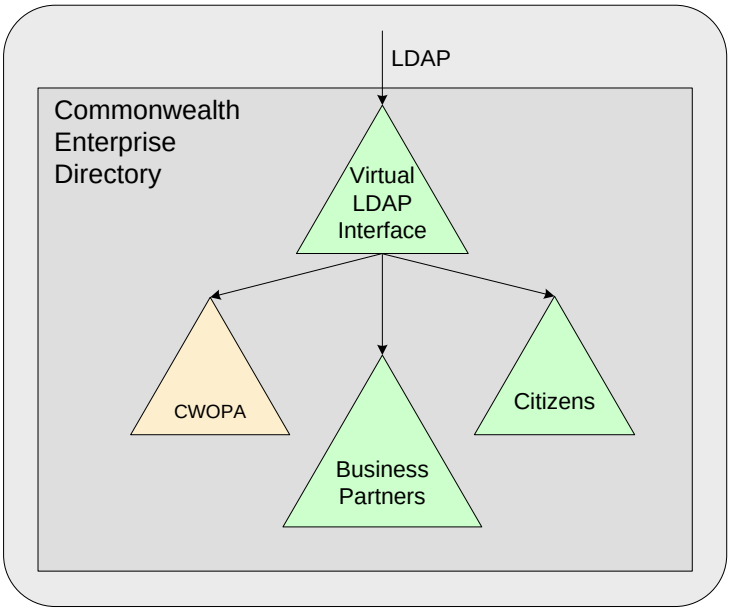

Figure 1 – CoPED Architecture

Supporting document GEN-SEC013B - *Directory Services Architecture* describes CoPED as a virtual directory service connecting three logical domains: *Employee*, *Business Partner*, and *Subscriber*. As shown in Figure 1, *CoPED Architecture,* CoPED is to host these three domains as three separate directory instances (where the *Employee* instance will actually be CWOPA for some period of time). However, for the initial deployment, and some interim period thereafter, the virtual directory service will point at specific agency directories to effectively "populate" the *Business Partner* and *Subscriber* domains of CoPED, as shown in Figure 2 – *Initial CoPED Architecture*.

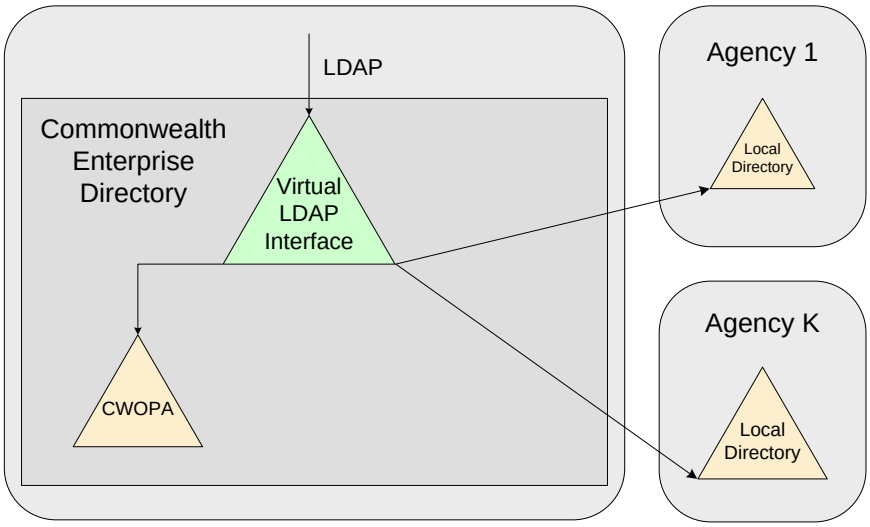

Figure 2 – Initial CoPED Architecture

3.4 Virtual Directory

Virtual directories provide a customizable view of all or part of one or more identity stores, and include features such as caching, filtering, data transformations, and schema transformations. For the CoPED service diagrammed above, the virtual directory would include:

- Two servers operating as redundant virtual directory engines, which:
  - Provide administration tools to define the configuration.
  - Interact with a configuration store to save and retrieve the configuration.
- The virtual directory servers to provide the following features as appropriate:
  - Cache directory data to improve performance.
  - Filter the directory tree so that only appropriate objects are published (for example, filter out non-employees so that only employees from CWOPA are seen in the *Employee* directory branch).
  - Filter the schema to only publish appropriate attributes (for example, if the source identity store includes SSN, do not publish SSN in CoPED).
  - Transform the published data (such as set a value to the Proofing Level attribute based on which identity store sourced the user).
  - Transform the published schema (for example, read the 8-digit driver's license ID as a numeric, but publish it as an 8-character alphanumeric).
- Back-end access to source identity stores via Lightweight Directory Access Protocol (LDAP) or SQL/ODBC.

4. Agency Master Directory Implementations

As discussed in Section 2, the agencies will be able to leverage the commonwealth's investment in enterprise directory technology by using it to develop their own master directories. An agency's master directory represents its own internal array of identity stores of non-employee users in a single, multifunctional view. Using these master directories to synchronize with CoPED

will simplify connectivity as well as provide the agencies with a consolidated view of their own identity information for internal application development and reporting.

The agencies can leverage either the virtual directory technology described above or the metadirectory technology described in Section 4 of supporting document GEN-SEC013B - *Directory Services Architecture*, for building their master directories. Each agency will also have the option of creating two directories (one for each of the *Business Partner* and *Subscriber* domains), or a single directory with an attribute or group that can definitively determine to which domain each user belongs.

Although the CoPED architecture calls for each agency to have a master directory, most agencies will not have these master directories in place at the beginning of the IPAM Architecture deployment. As an interim step, the CoPED virtual directory server will be able to leverage individual identity stores from among the agencies to publish those entities for the Shared Authentication Service. This is described in Section 5.1.

The metadirectory service to be included in the CoPED architecture is described in Section 4 of GEN-SEC013B - *Directory Services Architecture*.

- **CoPED RDN** – For the overall CoPED virtual directory interface, the CoPED GUID will form the user's Relative Distinguished Name (RDN), which is the user's unique ID within the bottom-most container. Since CoPED GUID will always be unique across all three namespaces, it will also always be unique within any single container. And since it will always be invariable (no matter what other changes a user undergoes), it makes an excellent choice as the base key for the directory.

- **Business Partner and Subscriber RDN** – For the same reasons as above for using CoPED GUID for the overall CoPED RDN, the CoPED GUID will also be used as the RDN in the new directories for both the Business Partner and Subscriber domains. Since the Employee domain will leverage the existing CWOPA Active Directory forest, that directory will continue to use the Active Directory (AD) default of Common Name (cn) as its RDN.

- **Synchronization Link** – As described in Section 4 of supporting document GEN-SEC013B *Directory Services Architecture*, the CoPED GUID will be the unique ID that is keyed for connecting CoPED with each of the agency user identity stores (existing stores initially, eventually migrating to agency master directories).

5. Transitional Model

An intermediate model is recommended for agencies in the interim before their master directories are deployed. In this intermediate model, the CoPED virtual directory technology will incorporate certain specified agency identity stores in

place of (or in addition to) the CoPED *Business Partner* and *Subscriber* directories.

The architecture model defined in Section 2 is based on agencies creating their own master directories to support CoPED synchronization. Most agencies however will not create these master directories within the time frame required to initially deploy CoPED for the planned Shared Authentication Service. Also, the *Business Partner* and *Subscriber* directory servers will not likely be deployed in that time frame. Therefore, the initial CoPED deployment will incorporate an alternate model, the Community Model Hybrid to bridge the transition to the Shared Directory Services architecture.

5.1 Intermediate Step – Community Model Hybrid

Like the Distributed Model, the Community Model Hybrid utilizes separate logical directories for each named domain to simplify administration of the domains and simplify a staged rollout strategy. Unlike the Distributed Model, however, this model populates CoPED directly from specific agency sources, using the virtual directory to incorporate those sources and publish the users as part of CoPED, rather than limiting the virtual directory to *Business Partner* and *Subscriber* for non-employees (CWOPA would still be used for employees).

This intermediate step allows agencies to integrate with the IPAM shared services quickly without waiting for the creation of their master directories. It also isolates those applications accessing the shared directory and authentication services from the migration of the virtual directory's source (from the local agency identity stores to the *Business Partner* and *Subscriber* directory servers).

5.2 Migration Path

The actual migration from the initial deployment to the final architecture can follow several paths, due to the isolation provided by the virtual directory server. One recommended path is presented in the phases below; many of the steps could be reordered and still reach the same final goal. Reaching that goal is the important point, not the order of the steps.

5.2.1 Phase 1

The initial deployment would provide the virtual directory, which would only publish CoPED as a virtual view of employees gathered from CWOPA. The virtual directory would filter all non-employees in CWOPA out of the published view. For this initial deployment, those employees not in CWOPA would not be published (implying, for example, that they would not be able to authenticate via the Shared Authentication Service described in GEN-SEC013B - *Directory Services Architecture*).

5.2.2 Phase 1a

Following directly after Phase 1, the immediate concern would be to publish non-employee users in CoPED. To provide access for users in the other

domains, the CoPED virtual directory would publish virtual views of select agency source identity stores for *Business Partner* and *Subscriber* domains. The agencies for this early deployment would be selected based on their readiness to participate in the IPAM architecture. Ideally, at this time both domains would be represented by local identity stores.

Although described here as a single phase, this process may consist of several discrete projects, each incorporating only a few local identity stores into the virtual directory views. These projects would be independent of each other, and could be carried out in serial, in parallel (if manageable with the resource constraints of the commonwealth), or partially overlapping.

5.2.3 Phase 2

At some point after the initial architecture is deployed, the *Business Partner* directory server will be added, and the link from the virtual directory would transition to that directory from the local identity stores that had been used. For this step to happen, the identity data from those local identity stores are to be populated into the *Business Partner* directory, requiring the deployment of the metadirectory service with connections to at least the *Business Partner* directory and the current local identity stores providing business partner users. See Section 4 of GEN-SEC013B - *Directory Services Architecture,* for more details about the metadirectory service.

Once the transition is complete, the virtual directory would no longer publish business partner users from local identity stores. Although described here as a single phase, this process would also consist of several discrete projects, including the deployment of the *Business Partner* directory and deployment of the metadirectory service and its connection to those local identity stores already incorporated into the architecture. Although these would be separate projects, they would have dependencies requiring them to be performed in the order noted (deploy *Business Partner* directory, deploy metadirectory service, and connect metadirectory to local identity stores). The final step would be to re-point the virtual directory away from the local identity stores to the new *Business Partner* directory.

5.2.4 Phase 3

In the final Phase, the *Subscriber* directory server would be added, and the link from the virtual directory would transition to that directory from the local identity stores that had been used. For this step to happen, the identity data from those local identity stores needs to be populated into the *Subscriber* directory, requiring the addition of metadirectory connections to the *Subscriber* directory and the remaining local identity stores providing citizen users. See Section 4 of GEN-SEC013B - *Directory Services Architecture,* for more details about the metadirectory service.

Once the transition is complete, the virtual directory would no longer publish any users from local identity stores. Although described here as a single phase, this process would also consist of several discrete projects, including

the deployment of the *Subscriber* directory and connection of the metadirectory to those local identity stores already incorporated into the architecture. Although these would be separate projects, they would have dependencies requiring them to be performed in the order noted (deploy *Subscriber* directory, connect metadirectory to local identity stores). The final step would be to re-point the virtual directory away from the local identity stores to the new *Subscriber* directory.

6. Directory Architecture

Once the CoPED model has been determined as in Section 2, the detailed directory architecture can be defined. This section provides that definition.

6.1 Domain Selection

The different domains (*Employee*, *Business Partner*, and *Subscriber*) not only represent different types of identities but more importantly represent them for different purposes. Therefore, they do not need to follow the same architecture internally. This section defines the namespace architecture to be implemented for each domain.

6.2 Suffix Selection

The suffix or root of a directory defines the base container for the directory server. Since administrators and users are accustomed to the CWOPA suffix, the CoPED suffix is to continue to be *dc=PA, dc=LCL* (note that CWOPA is actually an Organizational Unit, or *ou*, under that suffix).

The virtual directory service allows great flexibility in selecting suffixes for each of the underlying directories. For transparency and ease of integration, however, each of the new Enterprise domain directories (*Business Partner* and *Subscriber*, and in the future possibly *Employee*) is to use the same suffix: *dc=PA, dc=LCL*.

6.2.1 Directory Tree Models

Under that common suffix, however, each domain will have a Directory Information Tree (DIT) specifically customized to the needs of the applications that will access it. These are detailed separately below.

Since user objects need to remain in the directory even after they are no longer active, there will be a state where the object exists but is disabled and marked as an archive. This may be implemented as a separate archive branch (Organizational Unit) within each directory where users are moved, or by attributes or groups that will ensure access is denied if the associated user attempts to authenticate, or by equivalent methods determined during the detailed implementation design. This archive state ensures that deactivated users can be referenced when tracing historical information, especially audit trails and other forensics. By maintaining the archived users within each domain's directory, the directory server's internal functionality for forcing ID uniqueness can be leveraged to ensure that CoPED GUIDs

assigned to that directory are unique forever, and are never reassigned. The CoPED GUID is discussed thoroughly in supporting document GEN-SEC013B, *Directory Services Architecture*.

6.2.1.1 Employee

The *Employee* domain will use CWOPA, which already has a rich, hierarchical DIT under the Organizational Unit (ou):

ou=CWOPA, dc=PA, dc=LCL

This is useful for CWOPA, enabling delegated administration across the agencies and bureaus and meeting other important Windows and Exchange needs. Since this exists, there is no reason to change it. The Relative Distinguished Name (RDN), which identifies a user within a container, is defined in Active Directory as the user's Common Name (cn).

6.2.1.2 Subscriber

The *Subscriber* domain is very simple. Since each Subscriber can have interactions with several agencies and applications, there is no one application that can be considered to "own" that user's identity object, so all objects can be contained in a single flat OU:

ou=people, ou=Subscriber, dc=PA, dc=LCL

Within that container, each user's RDN is to be the CoPED GUID. In some cases, the selected product will not efficiently handle OUs as large as the *Subscriber* directory could be (potentially over twelve million with Real ID). In that case, adopt a design that utilizes efficient techniques for dividing the *ou=people* container into more manageable pieces. One simple, recommended process would be to break the numeric part of the CoPED GUID into blocks according to the size recommended for the directory server product and put each block into its own OU. Although it might seem useful to divide the *Subscriber* logically (for example, by counties), that will not provide the desired simple benefit, as then a process will need to be created to move users to new OUs when the users move between counties, and any references to the user's Distinguished Name (DN) that were stored by any application in the commonwealth will need to be informed that the user's DN was changed. Plus, if a county grew too large, the OU would have to be divided again.

If the archive is created as a separate branch, it would be a second flat OU:

ou=archive, ou=Subscriber, dc=PA, dc=LCL

If the *people* OU gets broken up as described above, the OU division used for *people* is to be replicated in the *archive* OU.

6.2.1.3 Business Partner

Unfortunately, unlike the *Subscriber* directory, the *Business Partner* directory domain can be complex. However, because each Business Partner user can fulfill multiple roles for differing agencies and applications, no one organization can be considered to "own" a user's identity object. Therefore, like *Subscriber*, all objects can be contained in a single flat Organizational Unit:

ou=people, ou=BusinessPartner, dc=PA, dc=LCL

Within the container, each user's RDN is to be the CoPED GUID.

In some cases, the selected product will not efficiently handle OUs as large as the active *Business Partner* directory could be (estimated at a few hundred thousand). In that case, adopt a design that specifies efficient techniques for dividing the *ou=people* into more manageable pieces. As with *Subscriber*, a simple process is to be used, such as simply breaking the numeric part of the CoPED GUID into blocks according to the size recommended for the directory server product, and putting each block into its own OU.

If the archive is created as a separate branch, it would be a second flat OU:

ou=archive, ou=BusinessPartner, dc=PA, dc=LCL

If the *people* OU gets broken up as described above, the OU division used for *people* is to be replicated in the *archive* OU.

6.3 Schema Architecture

The objects selected for initial CoPED use are listed in GEN-SEC014F *CoPED Schema.* The schema architecture is based on the IETF standard object class *inetOrgPerson*, which defines the bulk of the attributes used for CoPED. These attributes are specified in RFC 2798. While *inetOrgPerson* provides most of the required attributes, the architecture does use additional attributes. These attributes are to be defined as belonging to custom object classes that are structural under *inetOrgPerson*.

Each of the three namespace directories is to include one initial custom object class. Based on the naming conventions defined in Section 6.3.3, these would be *copedEmplPerson*, *copedBusPartPerson*, and *copedSubscriberPerson*. If additional custom object classes are needed, the governance process defined in Section 6.3.1, *Attribute Selection Policy*, provides for their creation. Custom object classes would group new custom attributes defined for a specific purpose that are not applicable to all users within a given domain. For example, a custom object class might be created called *copedPIVcardPerson* to hold custom attributes related to a PIV card that don't apply to all business partner users.

In general, only a minimum set of attributes is to be included in CoPED. The inclusion of personally identifiable attributes is to be reviewed for compliance with existing laws, regulations, and standards governing their use. This limitation, however, only means that such attributes (if belonging to a standard object class) are not to be populated with data. Standard attributes are not to be removed from *inetOrgPerson* or any of its superiors.

6.3.1 Attribute Selection Policy

This section defines the governance model for managing the schema. As stated in Section 7, *Governance and Administration* of GEN-SEC013B, the primary governance for CoPED rests with Enterprise Architecture. This includes setting and modifying policy as needed, issuing exceptions to those policies when appropriate, and enforcing the policies.

Schema governance is to flow from the strategic directory usage outlined in Section 2, *CoPED Directory Services*, of GEN-SEC013B, which states that CoPED serves as the repository for users' authentication credentials for the Shared Authentication Service, and stores those additional information attributes that are shared across many agencies. Consider legal compliance as the next guiding principle, especially to meet the privacy requirements of existing laws, regulations, and standards governing the use of personally identifiable attributes.

Additional policies may be added by Enterprise Architecture as circumstances dictate. Policies do not need to be created or modified to encompass every requested exception; an exception can be approved and implemented as an exception without changing the underlying policy.

Based on those strategic goals and guiding principles, an appropriate initial set of attributes has been established and published in GEN-SEC014F - *CoPED Schema*. Each new attribute is to be documented in a table format similar to the SEC014F spreadsheet, which includes the name, description, data type and format, sample syntax, the primary object class the attribute belongs to (could be more than one), whether the attribute is mandatory for that object class, and whether it is single or multi-valued. Additional information may be provided if useful, such as indexing, data for complex attributes, and access permissions if different from the object as a whole.

While many of these attributes already exist for employees in the CWOPA schema, there are several that do not exist and will need to be added as custom attributes in the same manner as for the other CoPED underlying directories.

6.3.2 Object Identifiers

Whenever possible, CWOPA is to assign standard Object Identifiers (OID numbers) to all of the custom object classes and attributes. These unique OID numbers are to be assigned respective to the CoPED registered private

enterprise numbering space, as defined in Section 3 of GEN-SEC013B - *Directory Services Architecture*.

6.3.3 Naming Conventions

Any custom attributes or object classes that are added are to be named in accordance with the following conventions. All such custom names are to begin with the prefix *coped* (all lower case) to indicate they are added for CoPED. They are to then have as-meaningful-as-possible names appended, with the first letter of each word capitalized and all words run together. For example, the custom object classes for the three domains will be: *copedEmplPerson, copedBusPartPerson*, and *copedSubscriberPerson*. As seen, structural object classes are to also give an indication of their superiors.

6.4 Directory Infrastructure

The following three subsections describe the architectural infrastructure for CoPED.

6.4.1 Physical Topology

As described in GEN-SEC013B - *Directory Services Architecture*, CoPED comprises a virtual directory server creating an enterprise view of three distinct namespaces, which may be implemented as separate directory instances or even physically separate directory servers. Best practices suggest two redundant virtual directory servers, providing both load balancing and failover.

The first namespace, *Employee*, will use the existing CWOPA Active Directory forest and thus will require nothing more than access to two CWOPA domain controllers (DCs). For performance considerations, at least the primary DC is to be dedicated to CoPED; the second DC could be treated by the virtual directory server as a failover and may be shared with other uses, although if feasible it is to also be dedicated to CoPED.

The other two namespaces, *Business Partner* and *Subscriber*, will be implemented initially as separate instances of a single directory service, again on two redundant servers to provide load balancing and failover. As usage grows, these instances can be separated out to individual servers with almost no impact on the configuration, and the number of servers can be increased as well to provide more capability as needed.

Due to the nature of the virtual server technology, the ideal physical topology is to have all of the servers co-located on a high speed backbone in the Data Tier, including:

- Virtual directory servers
- AD DCs
- Business Partner and Subscriber directory servers

In addition, the Enterprise Web Access Management servers (Policy Server and Federation Server) belong in the Application Tier, as close as possible to the Data Tier backbone for expedited access to CoPED.

Although not required, it is also desirable for any agency identity store (being leveraged by the virtual directory server in the initial stages following the Community Model Hybrid) to be co-located in the Data Tier with CoPED, or as close as possible. User authentications where the identity is coming from a distant store will experience noticeable delays.

6.4.2 Replication

This infrastructure is to adhere to existing commonwealth protocols and policy for mission critical data systems. As such, each of the required servers comprising CoPED (virtual directory server and *Business Partner* and *Subscriber* directory servers) will be redundant to meet the appropriate availability and performance needs. As noted above, however, these systems will be co-located and therefore no directory replication will be needed for geographical distribution. The only replication provided will be between the redundant systems and the normal AD replication internal to CWOPA.

6.4.3 Disaster Recovery

Adhere to existing Commonwealth protocols and policy for mission critical data systems. As noted above, the systems comprising CoPED are to be replicated to ensure availability in case of failure by any one component. In addition, backups are to be made of appropriate data, and kept in a secure, off-site location to ensure recovery from a disaster affecting the CoPED location. Data to be backed up includes:

- *Business Partner* and *Subscriber* directory server configurations
- *Business Partner* and *Subscriber* directory server directory data
- Virtual directory server configuration

7. Security Issues

The security for any government directory warrants careful examination. This is especially true for CoPED because it is a core component with the access control systems. The CoPED system configuration and encryption are to adhere to existing Commonwealth protocols and policy for securing mission critical data systems, including restricting physical access to them. Refer to supporting document GEN-SEC013B for a discussion of these issues and corresponding policy standards.

7.1 Encryption Guidelines

Private and sensitive data is to be encrypted in place, with an encryption level meeting Commonwealth protocols and policy for securing mission critical and sensitive data as well as the 3DES or AES encryption standard.

When initially deployed, the following CoPED attributes are to be encrypted in place:

- *userPassword* -- depending on directory server product, probably stored as encrypted one-way hash rather than merely encrypted password
- Certain agency unique IDs, if specified by the agency
- When data is synchronized from agency identity stores, the agency may specify attributes to be encrypted

7.2 Web Services (WS-Security)

The OASIS Web Services Security Technical Committee has released, and OASIS has approved version 1.1 of the *Web Services Security: SOAP Message Security* specification, dated Feb. 1, 2006. This specification describes enhancements to SOAP messaging to provide message integrity and confidentiality. The specified mechanisms can be used to accommodate a wide variety of security models and encryption technologies.

While there is no immediate need to implement WS-Security features for Web Services communications over Simple Object Access Protocol (SOAP), a Service Oriented Architecture (SOA) is a Commonwealth goal and the CoPED implementation is not to be in a manner that might preclude their later inclusion. The commonwealth (and the Governor's Office of Administration, Office for Information Technology, Enterprise Information Security Office in particular) will continue monitoring Web Services developments for future incorporation into CoPED.

8. Virtual Directory Guidance

Because a virtual directory server does not store directory data per se, many of the standard guidelines that apply to typical directories (or databases) do not apply. This section discusses some of the general guidelines that do apply to virtual directory technologies; many other detailed implementation guidelines vary depending on the product selected.

8.1 Identity Data Synchronization Architecture

The Identity Data Synchronization Architecture is used to populate the CoPED as well as to synchronize user data among CoPED and the various agencies identity stores. This synchronization is performed by metadirectory technology. A metadirectory is a multidirectional synchronization product capable of interfacing with a variety of identity repositories.

The master directory serves as the agency own personal "Enterprise Directory" by publishing all user objects and the subset of their attributes that are to be visible outside the agency. A metadirectory "publishes" identity data to a directory or other identity store by creating user objects, and writing attribute values to user objects in those identity stores.

Each agency may leverage the same metadirectory technology employed for this synchronization service in a separate instance to aggregate its user data into the master directory, or may leverage the virtual directory technology as described in GEN-SEC013B - *Directory Services Architecture,* to virtually publish its users without aggregation.

As with most IT initiatives, the IPAM Identity Data Synchronization service is deployed in phases to provide a smooth transition, according to the Community Model Hybrid described below.

8.2 Scaling

Like most successful projects, CoPED will begin with a fairly small set of data and then grow. Ultimately, CoPED is projected to provide service for well in excess of twelve million users. At least three separate directory instances will be instantiated with its virtual directory. Key scaling guidelines to take into consideration include:

- **Memory** – Virtual directory technology has one major drawback. By pointing to (rather than aggregating) individual attribute data, the virtual directory inserts an extra network hop (and its resulting delay) when publishing data. The vendors mitigate this drawback by caching some part of the directory data, typically in RAM. By providing the virtual directory server with the largest amount of RAM feasible, CWOPA will increase the likelihood of any given piece of data being found in the cache and minimize access delays.

- **Load Balancing** – By adding redundant virtual directory servers, the number of simultaneous accesses can be increased as usage increases. This guideline corresponds with those of other technologies.

8.3 Performance

Similar to scaling, key performance guidelines include:

- **Location** – CWOPA can also minimize the inherent virtual directory delay by locating the source identity stores (or at least a consumer replica) on the same high-speed backbone as the virtual directory server. This ensures that the delay due to the additional hop is kept negligibly small. For example, in the initial architecture, a CWOPA Domain Controller will be co-located in the Data Tier with CoPED.

- **Indexing** – When accessing data from the source identity stores, the virtual directory will identify objects or records to access. Access speed is best when the most useful key is stored by the virtual directory to identify the record to access, and all such keys are well indexed in the local identity stores.

- **Cache Management** – When the virtual directory server cannot access a source identity store, it is to still respond to the LDAP request. In most cases, the ideal is for the server to respond from its cache if the requested data is already there, even when the source is not reachable. In some cases, however, the commonwealth may not want to allow any access if the source data is unreachable. This is configurable by the commonwealth.

- **Load Balancing** – Redundant, load balanced virtual directory servers reduce the load on any one server, improving performance for all. This guideline corresponds with those of other technologies.

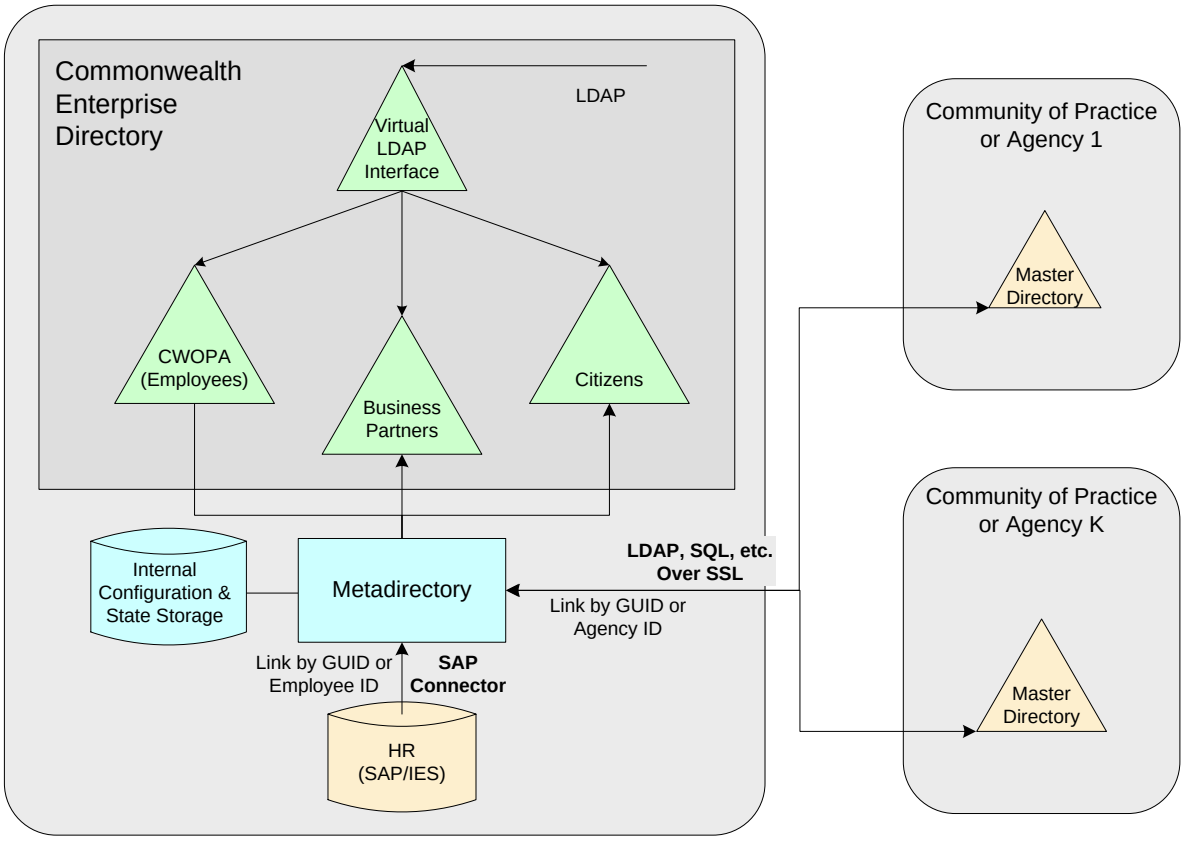

Figure 3 – Identity Data Synchronization Architecture

The following subsections describe the phases for that transition.

8.3.1 Phase 1

The first phase of the transition focuses on the *Employee* domain. It requires the installation of the metadirectory's synchronization engine and setup and definition of its internal data store (sometimes called a metaverse). A connection from the SAP HR Directory (IES) serves as the authoritative source for employees and their various CoPED attributes. A connection to CWOPA using Active Directory Services Interface (ADSI) will serve as the *Employee* underlying domain directory.

As shown in *Figure 4 – Phase 1 Synchronization Architecture*, Phase 1 configures the metadirectory to:

- Read employee records from IES (filtering out any non-employee records encountered).

- Verify that the user does not already exist as an employee (such as checking employee ID, name matches). Flag any potential matches as errors to be reviewed manually by the metadirectory administrators.

- Create a new CoPED GUID for the user and return it to IES, which may or may not store the value.

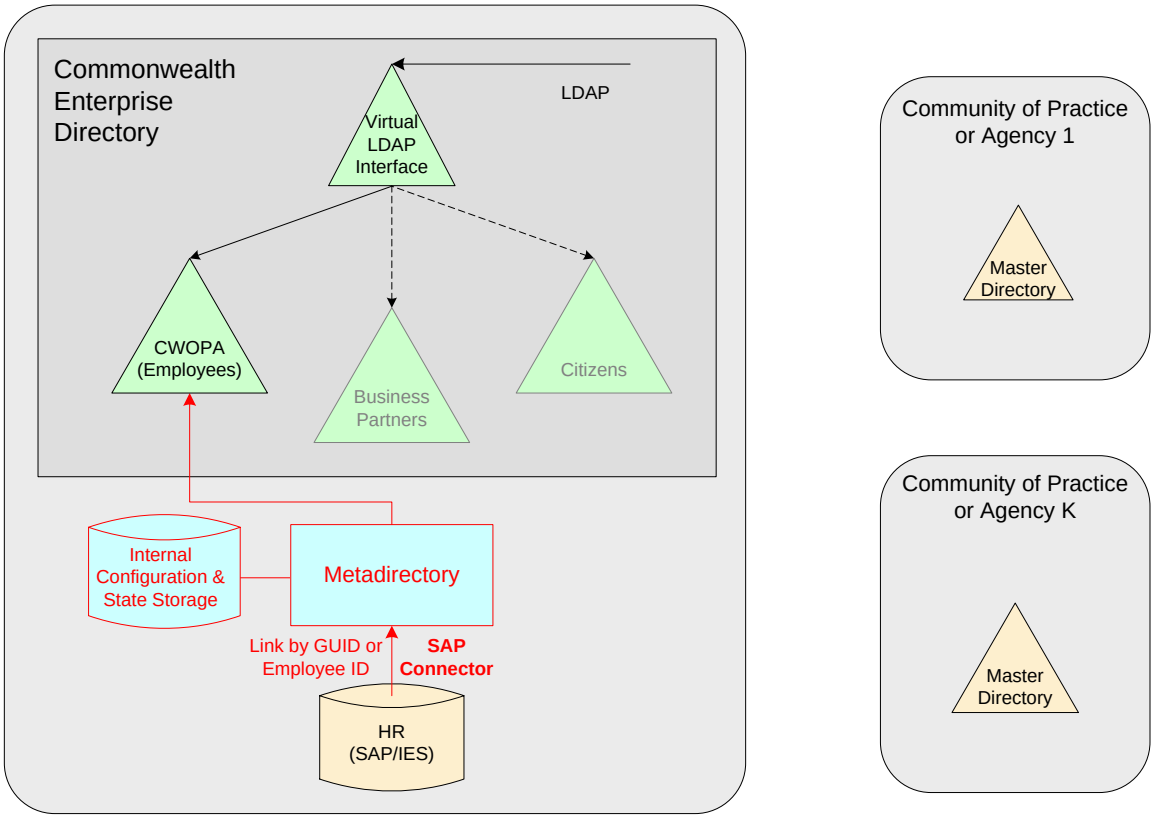

Figure 4 – Phase 1 Synchronization Architecture

Store the important user data in its internal data store, including both the CoPED GUID and the IES key (employee ID).

- Write each new user object to CWOPA.

- Write updated attributes values to CWOPA and back to IES as appropriate.

This populates employees and their attributes to CWOPA as required to support the Shared Authentication Service defined in GEN-SEC013C - *Access Management and Control*.

8.3.2 Phase 2

The second phase of the transition focuses on the *Business Partner* domain, beginning with First Responders. It builds on the deployed metadirectory

service by adding connectors to appropriate agency master directories to discover new business partner user objects and collect appropriate identity data from them. This phase also adds a connector to the *Business Partner* underlying domain directory within CoPED, to publish the business partner users into that directory. The Phase 2 connectors use the standard access protocols for the connected directories, most likely LDAP (or ADSI for Active Directory).

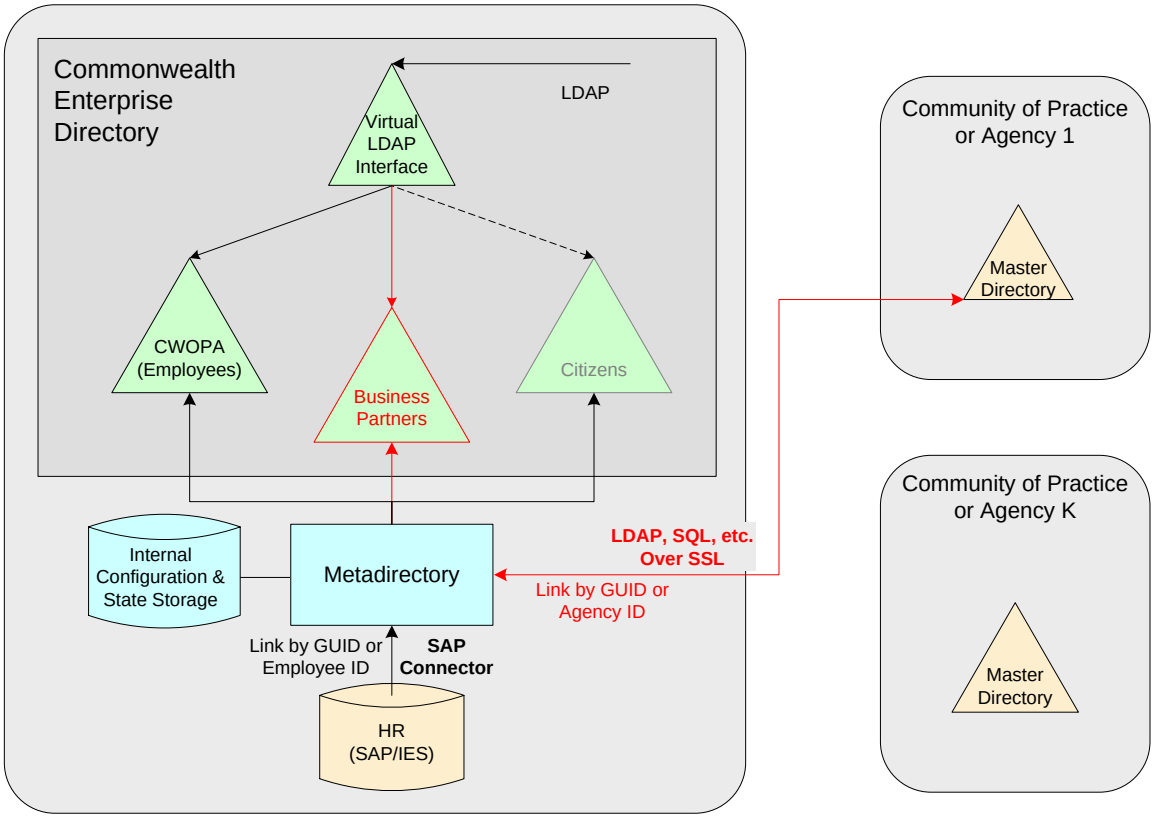

Figure 5 – Phase 2 Synchronization Architecture

As shown in Figure 5 – Phase 2 Synchronization Architecture, Phase 2 configures the metadirectory to:

- Read business partner user objects or records from one or more agency identity stores that are authoritative for those users.
- Verify that the user does not already exist as a business partner user (such as checking name matches, local key/identifier). Flag any potential matches as errors to be reviewed manually by the metadirectory administrators.
- Create a new CoPED GUID for the user and return it to the connected identity store, which may or may not store the value.

- Store the important user data in its internal data store, including both the CoPED GUID and the local key (such as PIV card serial number or other identifier used by that identity store).

- Write each new user object to the *Business Partner* underlying domain directory within CoPED.

- Write out updated attributes values as appropriate, both to the domain directory and to connected master directories.

This populates business partner users and their attributes to the *Business Partner* underlying domain directory, as required to support the Shared Authentication Service defined in GEN-SEC013C.

8.3.3 Phase 3

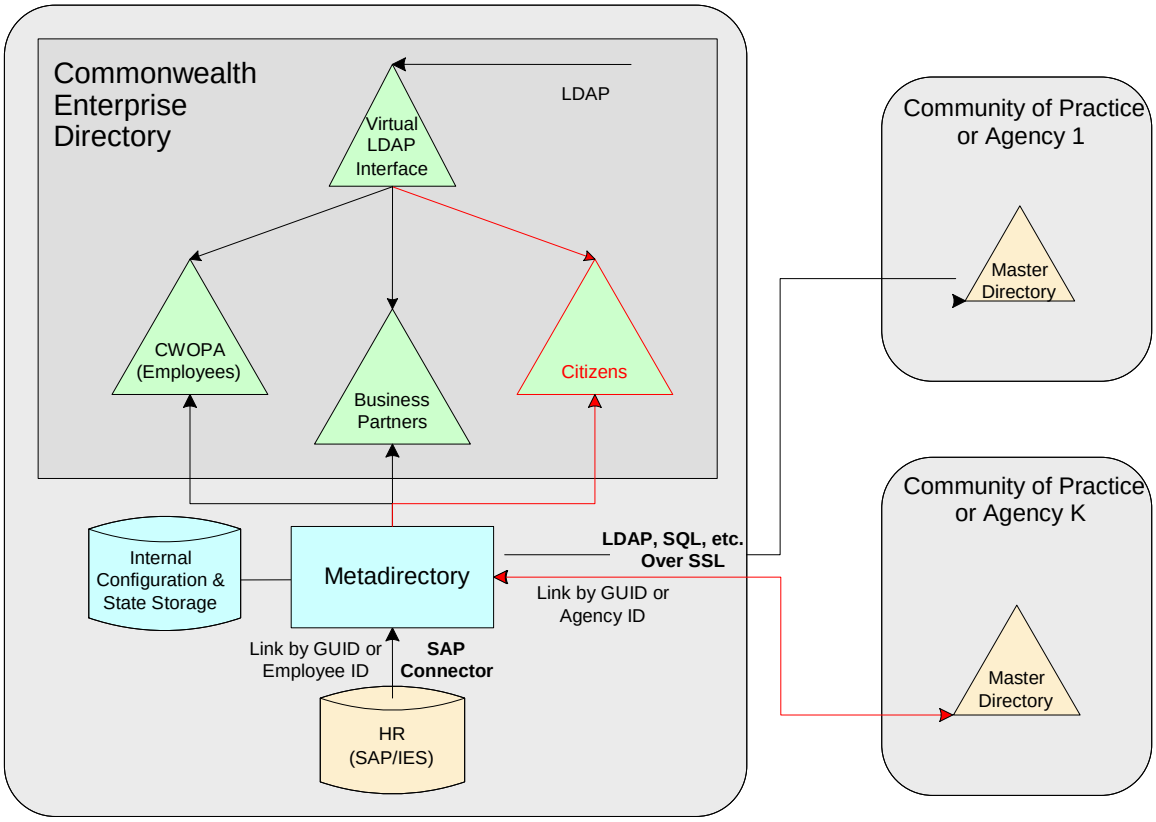

Figure 6 – Phase 3 Synchronization Architecture

The final phase of the transition focuses on the *Subscriber* domain. It builds on the deployed metadirectory service by adding connectors to appropriate agency master directories to discover new Subscriber user objects and collect appropriate identity data from them. This phase also adds a connector to the *Subscriber* underlying domain directory within CoPED, to publish the Subscriber users. The Phase 3 connectors use the standard access protocols for the connected directories, most likely LDAP (or ADSI for Active Directory).

As shown in Figure 6 – *Phase 3 Synchronization Architecture*, Phase 3 configures the metadirectory to:

- Read Subscriber user objects or records from one or more agencies identity stores, which are authoritative for those users.
- Verify that the user does not already exist as a Subscriber user (checking name matches, local key/identifier). Flag any potential matches as errors to be reviewed manually by the metadirectory administrators.
- Create a new CoPED GUID for the user and return it to the connected identity store, which may or may not store the value.
- Store the important user data in its internal data store, including both the CoPED GUID and the local key (such as PA Driver's License number or other identifier used by the local identity store).
- Write each new user object to the *Subscriber* underlying domain directory within CoPED.
- Write out updated attribute values as appropriate, both to the domain directory and to connected master directories.

This populates Subscriber users and their attributes to the *Subscriber* underlying domain directory, as required to support the Shared Authentication Service defined in GEN-SEC013C.

8.4 Standards

The Identity Data Synchronization service follows all standards listed in ITP-SEC013 - *IPAM Architectural Standard-Identity Management Services*, as appropriate. In particular, all access to CoPED (to publish users or attribute modification or gathering) uses the Lightweight Directory Access Protocol v3. Access to the connected agency master directories use the standard access protocol most appropriate for each store, which is ADSI for Active Directory, LDAP v3 for other directories, or SQL for databases.

9. Related ITPs/Other References

- GEN-SEC013B - *Directory Services Architecture*
- GEN-SEC013C - *Access Management and Control*
- GEN-SEC013G - *Public Key Infrastructure*

10. Authority

- Executive Order 2011-05, Enterprise Information Technology Governance

11. Publication Version Control

It is the user's responsibility to ensure they have the latest version of this publication. Questions regarding this publication are to be directed to RA-itcentral@pa.gov.

This chart contains a history of this publication's revisions:

Version	Date	Purpose of Revision
Original	9/7/2006	Base Policy
Revision	9/25/2009	Updated Format
	4/2/2014	ITP Reformat