

Virtual Node

Integrated Project Team

 Guidance and Recommendations Document

Version: 1.0

Revision Date: January 29, 2013

# Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Version Number | Description of Change | ChangeEffective Date | ChangeEntered By |
| 0.3 | Initial draft delivered to VN IPT | 12/17/2012 | EPA |
| 1.0 | Revised to reflect comments from VN IPT members. | 1/29/2013 | EPA |

Table of Contents

[Revision History ii](#_Toc343543341)

[1 Introduction 1](#_Toc343543342)

[1.1 Virtual Node Background 1](#_Toc343543343)

[2 Virtual Node Context and Drivers 2](#_Toc343543344)

[2.1 Anticipated Benefits 2](#_Toc343543345)

[2.2 Use Cases 3](#_Toc343543346)

[3 Virtual Node Implementation Patterns 4](#_Toc343543347)

[3.1 Key Components in Virtual Node Implementation Pattern Options 4](#_Toc343543348)

[3.2 Virtual Node Accesses Partner Data Store 5](#_Toc343543349)

[3.2.1 Implementation of Local Trading Partner Managed Staging Tables 7](#_Toc343543350)

[3.3 Virtual Node Accesses Shared Standardized Staging Tables 7](#_Toc343543351)

[3.3.1 Implementation of Standardized Virtual Node Managed Staging Tables 9](#_Toc343543352)

[3.4 Virtual Node XML Mapping Component 9](#_Toc343543353)

[4 Virtual Node Communication Model 10](#_Toc343543354)

[5 Virtual Node Administration 11](#_Toc343543355)

[5.1 Virtual Node Instance Administration 12](#_Toc343543356)

[5.1.1 Virtual Node Administration Portal 13](#_Toc343543357)

[5.2 Virtual Node Security 13](#_Toc343543358)

[5.2.1 Virtual Node Users and Roles 13](#_Toc343543359)

[5.2.2 Virtual Node Security Module 14](#_Toc343543360)

[6 Implementation and Operational Recommendations 14](#_Toc343543361)

[6.1 Transitioning to the Virtual Node 14](#_Toc343543362)

[6.2 Resource Planning 15](#_Toc343543363)

[6.3 Help Desk Requirements 16](#_Toc343543364)

[6.4 Training and Documentation 17](#_Toc343543365)

[6.5 Change Management and Governance Considerations 18](#_Toc343543366)

[6.6 Memorandum of Understanding Recommendations 19](#_Toc343543367)

# Introduction

This document provides recommendations related to the implementation and usage of a cloud-based Virtual Node for Trading Partners on the Exchange Network. The content in this document is based on input from the Exchange Network Virtual Node Integrated Project Team (IPT)[[1]](#footnote-1) that was formed to discuss and investigate the adoption of a Virtual node platform on the Exchange Network.

## Virtual Node Background

For many Exchange Network partners, maintaining a presence on the Exchange Network (EN) means implementing and maintaining an Exchange Network Node. To date, the Exchange Network has used a federated approach to Node implementation with each partner maintaining a local instance of Node software that complies with a core set of functionality as defined in the EN Protocol and Specification. This model alone may no longer be sufficient for meeting the needs of Network partners. Budget shortfalls, staff turnover, and changes in technology and program requirements can pose challenges to some organizations that wish to operate a Node on the Network. The maturity of cloud-based technologies may offer a path to cheaper and more efficient ways of managing infrastructure. A shared environment may also offer opportunities to help partners more easily meet their goals around data publishing.

EPA has researched cloud models for the Exchange Network and has prototyped a multi-tenant application model called the Virtual Node, which is analogous to a central shared node platform that can be individually configured by each partner to handle its different data flows. The purpose of this prototype was to assist the IPT and other participants in preparing requirements for a production-ready Virtual Node. EPA anticipates establishing and maintaining a Virtual Node infrastructure in a central shared cloud environment. This model was in response to existing and potential small and medium sized Exchange Network partners who would likely receive the most value and establish more cost efficient ways to manage their nodes and decrease costs. However, this does not preclude a Network partner from benefitting from the Virtual Node if it improves efficiency and experience based on their unique requirements. This new approach has potential to significantly simplify development and maintenance using inheritance features and plug-in support.

In the FY 2013 Exchange Network Grant Solicitation Notice, EPA defined the Virtual Node in the following terms: A Virtual Node is a central node server that can host any number of partner nodes: state, tribe, region, agency, etc. Each partner node is simply configured using wizards and forms on the central virtual node cloud server. This eliminates the need for maintaining a local node completely. Each partner node functions like a conventional node, only it is much simpler and less expensive to set up and maintain.

# Virtual Node Context and Drivers

## Anticipated Benefits

Through the course of the IPT activities, IPT team members discussed and agreed to a number of key anticipated benefits that Virtual Node Trading Partners could expect to receive as a result of adopting the Virtual Node platform. These benefits are documented below.

* **Minimize Operational Costs/Burden –** Implementation of the Virtual Node platform will allow Trading Partners to realize costs savings and burden reduction for day to day operations related to the maintenance of a node infrastructure and node software as this maintenance will be provided by EPA through the shared Virtual Node environment.
* **Reduce initial barriers and burden to EN participation –** New Trading Partners or Trading Partners that are interesting in expanding their presence on the EN will face fewer barriers to doing so. The initial upfront investment for the infrastructure, software, and technical resources familiar with the node specification will be reduced as a result of adopting the Virtual Node platform. This will allow for continued expansion of new Trading Partners on the network at less cost to those Trading Partners.
* **Simplify adoption and implementation of new data flows –** New data flows established by EN Trading Partners will be able to create shared components (e.g. XML style sheet mapping, data flow patterns/orchestration) that will be reusable across Trading Partners. As a result, Trading Partners will be able to leverage these resources. This model has proven to be successful with existing node implementations for Trading Partners and is thus a critical aspect of how the Virtual Node will be architected.
* **Enable data publishing strategy –** The Virtual Node platform will include capabilities that will simplify the requirements for publishing Trading Partner data, aligning with the Phase 2 goals of the Exchange Network.
* **Focused Exchange Network Investment –** Reduced burden and operational costs, and a simplified data exchange model with a centralized and shared Node architecture will allow Trading Partners to focus investment on mission oriented, business process focused tasks that the Exchange Network enables.
* **Stay current with network and data flow changes –** As changes occur to network specifications related to node protocols, data flow changes, or changes to specifications related to integration with additional network services such as the Exchange Network Discovery Services (ENDS) or Network Authentication and Authorization Services (NAAS), these changes will be implemented and deployed to Trading Partner Virtual Node instances without requiring significant investment for the community. As a result, Trading Partners will be able to more easily stay current with network and data flow changes.
* **Achieving a more stable data exchange environment –** Historically, Trading Partners have sometimes found it difficult to dedicate resources or invest in resources with the subject matter expertise to help maintain the uptime or availability of either their EN node presence or specific data flows. The Virtual Node operation and maintenance model will allow for EPA resources to provide the scale and flexibility to help maintain a more stable presence on the network for Trading Partners.
* **Guaranteed security standards across all partners –** As security protocols and standards evolve on the EN, the implementation of these standards will be “pushed” to Virtual Node instances without the need for Trading Partner investment in the resources needed to implement the changes.
* **Emergence of vendor neutral solutions –** Exchange Network Trading Partners will be able to participate in data exchanges without requiring the adoption of a vendor specific solution.

## Use Cases

This section of this document describes example use cases for interested parties who may adopt the usage of the Virtual Node platform.

The initial implementation of the Virtual Node platform will conform to the Node Exchange Network Protocol v2.1 specifications and remain current with changes to these specifications over time[[2]](#footnote-2). The Virtual Node can be used by Exchange Network Trading Partners to orchestrate any set of standard Exchange Network Node transaction patterns. These include: (1) Submit data to other Trading Partners on the EN, (2) Publish data via EN Query and Solicit processes, (3) Publish data via VN REST service and (4) Receive data (i.e. host a Submit service) from external EN Trading Partners.

1. **Submit data to other Trading Partners on the EN:** This EN pattern involves a data flow where an EN Trading Partner needs to submit data to another Trading Partner. A standard example of this type of use case would be where a Trading Partner would adopt the Virtual Node to manage the process of one or more of the national system flows. Note that for EN Trading Partners that have adopted staging table and plug-in concepts, the architectural principles endorsed by the IPT allow for the staging tables used for these flows to be either located within the Trading Partners architecture or be managed as part of the Virtual Node architecture. Alternatively, the Virtual Node can be used in data exchanges between facilities and states, state to state, tribe to state or other data flow patterns.
2. **Publish data via EN Query and Solicit processes:** This EN pattern involves a data flow where an EN Trading Partner requests data via an EN Query or Solicit web service from an EN Trading Partner who is using the Virtual Node. Note that Virtual Node implementers can configure these data flows to connect to data stores within their architecture or to staging tables within the Virtual Node architecture.
3. **Publish data via Virtual Node REST services:** This EN pattern involves a data flow where an EN Trading Partner or the public requests data via a REST service on a Virtual Node which is created automatically when a service is configured. Note that Virtual Node implementers can configure these data flows to connect to data stores within their architecture or to staging tables within the Virtual Node architecture.
4. **Receive data from external EN Trading Partners:** This EN pattern involves a data flow where an EN Trading Partner submits data via an EN Submit web service to an EN Trading Partner who is using the Virtual Node (i.e. the Virtual Node hosts a Submit service for the Virtual Node implementer). After the Virtual Node authenticates the requester and validates the payload, Virtual Node implementers can configure these data flows to connect to data stores within their architecture or to staging tables within the Virtual Node architecture in order to receive and stage the data for downstream business processes.

# Virtual Node Implementation Patterns

The Virtual Node platform allows for a flexible set of implementation patterns that can be adopted by Trading Partners. This section of the document discusses these implementation patterns.

A Trading Partner could choose to utilize a combination of the patterns described below. For example, a Trading Partner may want to adopt a data publishing strategy that utilizes a secure connection from the Virtual Node to a data store within the Trading Partner’s infrastructure but may want to utilize a Virtual Node staging table approach to facilitate the submission of data to a National System flow.

## Key Components in Virtual Node Implementation Pattern Options

The following key elements are identified in the diagrams found in Sections thru . These key components are used to depict the two primary implementation pattern recommendations for partners implementing the Virtual Node.

* **Virtual Node Boundary/Virtual Node:** The dotted line labeled Virtual Node Boundary delineates the components that would need to be managed by the EN Trading Partner (each described in detail below) as compared to the services configured by the EN Trading Partner within the Virtual Node infrastructure. The Virtual Node will meet all applicable EN Node specifications and include features such as role based security access, scheduling, and data flow configuration wizards.
* **Program System and Operational Data Store:** The Program System and Operational Data Store(s) represent the source system application and data storage of the Trading Partner. Trading Partners interested in making their data available on the Exchange Network (EN) for use cases that include EN Submits or for data flows related to data publishing (Query or Solicit) require a source data system(s).
* **View/Clone (Or) Staging Table(s):** In some cases, an EN Trading Partner may want to allow for a connection from the Virtual Node to a View/Clone of the Operational datastore to enable simple, agile, and/or rapid data access for data publishing purposes (rather than require data to be loaded into a staging area within the Virtual Node infrastructure). This component represents the portion of an EN Trading Partner’s architecture that the Virtual Node could connect to directly to provide data access.
* **XML Payload Generation (or) XML Mapping Engine:** These components represent the process by which XML payloads are either generated or inserted into the applicable datastore(s). For data publishing and traditional National System flows, this component would manage the file generation. In the event that an EN Trading Partner uses the VN to accept incoming transactions, this component would manage the receipt and loading of submissions into the applicable datastore.
* **Extract, Transform and Load (ETL) Processing** (Figure 2 only): In some cases, EN Trading Partners may have constraints that do not allow for direct access from the Virtual Node to local datastores within their DMZ. In these cases, an ETL process could be used to extract the appropriate data from the operational datastore(s) and load the data into Node Staging Tables that are part of the centrally maintained Virtual Node architecture. The ETL process would be created and maintained by the EN Trading Partner.
* **Virtual Node Staging Tables** (Figure 2 only): In cases where an EN Trading Partner requires Virtual Node Staging Tables, these can be hosted in the cloud/Virtual Node infrastructure. The Virtual Node would connect to these tables in order to fulfill EN requests.
* **External Trading Partner Infrastructure:** The Trading Partner in the diagrams could be any external Exchange Network Node presence or any Trading Partner application embedded with Node Client functionality (e.g., a web page, mobile app, etc.). In the case of an EN Submit or Solicit this component would most likely be an EN Node server that would be able to receive the Submit from the VN implementation. In the case of EN Query, the request could be generated by any application with the ability to generate and consume an EN Query service. In the event that an EN Trading Partner uses the Virtual Node to accept incoming transactions, this component could be either another EN Node server or an application that functions as an EN client and is submitting data to the EN Trading Partners instance of the Virtual Node.

## Virtual Node Accesses Partner Data Store

Figure 1 depicts an implementation pattern that utilizes the XML Mapping Engine component of the Virtual Node as described in Section . In this implementation pattern, Trading Partners are able to configure a database connection to a local data store that is within the Trading Partner’s infrastructure. This connection will be dependent upon the implementation of a secure connection between the Virtual Node and the Trading Partner’s environment (see Section 4).

The local data store that is used by the Trading Partner could be any data source that supports database connectivity and SQL statement extraction of the data. Example options for this local data store include:

* A data warehouse or data mart that the Trading Partner is utilizing for public report and data publishing
* Materialized Views of an operational data store
* A clone of an operational data store that is refreshed on a periodic basis
* A staging table or set of staging tables that the Trading Partner has implemented for a particular data flow (i.e. allowing for reuse of existing investments for Trading Partners who have adopted staging table based plug-ins)

Trading Partners will be able to configure multiple data source connections within their Virtual Node instance and connect to a variety of different data sources across multiple data flows.



Figure Virtual Node Accesses Partner Data Store

Key implementation notes for this option are noted in the bullets below.

* This implementation approach simplifies the number of components required to enable the exchange of data in that it may be possible to eliminate the Extract, Transform and Load (ETL) process that is required to move data from operational data stores to staging tables.
* Data that is made accessible in this approach will be as current as business rules allow. For example, in a case where a Trading Partner is using a View or Clone of an operational data store that affords a “real time” view to operational data, the data that is provided through the data exchange can be published or exchanged as soon as the constraints for a particular SQL statement are met. Meaning there would not be a dependency on a periodically scheduled process to ETL the data to a staging table model from which the data would be provided. This does not mean that data could be exchanged without the appropriate QA/QC process that is currently performed by EN Trading Partners. This QA/QC process can still be a prerequisite for data exchange as the SQL query that is configured in the Virtual Node could be structured to only retrieve data that has gone through this process.
* Trading Partners can choose the nature of the data source to connect to (e.g. DB copy in DMZ, staging tables, or operational datastore). This flexibility allows for reuse of existing EN investments by Trading Partners as well as affords the opportunity for Trading Partners to adopt new data publishing exchanges that may not be dependent on complex ETL and staging table architectural components.
* For existing data flows, data mappings may or may not remain intact. Due to the ability of the Virtual Node to connect to the data source of the Trading Partner’s choosing, each implementer will be able to reuse as much of their previous data mapping from the source database to the data exchange schema as they prefer.
* This implementation model will require a secure connection between the Virtual Node infrastructure and the Trading Partner’s infrastructure that will allow for database connectivity. Implementation options for this connectivity are discussed in Section 4. The VN IPT recognizes that this connectivity will require detailed coordination and discussion with appropriate security teams in order to ensure that that connectivity conforms to the Trading Partner’s security policies.
* This powerful new cloud model for secure connections between the Virtual Node cloud infrastructure and the Trading Partner’s infrastructure may be a challenge for Virtual Node adopters, but the simplicity and advantages make it well worth the trouble. Trading Partners interested in this type of connectivity model may need to address policies of this type within their organization in order to implement this pattern. EPA would like to explore this type of connection pattern with specific Trading Partners to see if new and emerging connectivity options can meet security requirements and policies and allow for this type of streamlined access to data for use cases like data publishing.

### Implementation of Local Trading Partner Managed Staging Tables

Trading Partners who are able to implement this simplified connection and data exchange model may still wish to utilize a staging table model in certain cases (e.g. previous investment has been made for an ETL process to use a set of staging tables). Use of the staging table model is still feasible in this implementation pattern as the data source that would be configured in the Virtual Node would be configured to access the staging tables within the local Trading Partner’s environment.

## Virtual Node Accesses Shared Standardized Staging Tables

Figure 2 depicts an implementation pattern that utilizes an ETL process implemented and configured by the Trading Partner that is used to move relevant data from a source system(s) within the Trading Partner’s environment to a staging table (or set of staging tables) in the Virtual Node infrastructure.

In this implementation pattern, Trading Partners will need to be able to configure a database connection from their local infrastructure to the Virtual Node staging table environment in order to implement a “push” model to provide the data to be exchanged via the Virtual Node. After data is pushed to the Virtual Node staging tables, data that is exchanged via a specific data flow would only come from the data that had been previously provided to the Virtual Node staging tables.



Figure 2 Virtual Node Accesses Shared Standardized Staging Tables

Key implementation notes for this option are noted in the bullets below.

* The ETL process to load the virtual node staging tables is expected to be accomplished by data transfers using existing technologies. This would potentially include bulk SQL operations from the state environment to the virtual node (e.g. Oracle SQL\*Net or SQL Server network traffic) or via secure file transfer of a database export or backup file and subsequent restoration by an EPA database administrator.
* Data that is made accessible in this approach may not be as current as business rules allow. In this model, Trading Partners will need to schedule or execute an ETL process on a recurring basis in order to move data ready for exchange from their local environment to the Virtual Node staging tables. The frequency of this ETL process will dictate how recent the data available to the data flow will be. This may not be ideal in uses cases where real time data access is needed for publishing or decision making.
* Trading Partners will need to design, develop, implement and maintain the ETL process that takes data from their operational data stores and transfers this data to the staging tables that are a part of the Virtual Node infrastructure.
* For existing data flows, data mappings may remain intact. In cases where Trading Partners have implemented staging tables for an existing data flow, it is possible that their existing ETL routines can be leveraged in order to move data from their local infrastructure to the Virtual Node staging table. This will be dependent on whether or not the staging table structure implemented in the Virtual Node environment is identical to the structure that is currently used by a given Trading Partner for a previously implemented data flow (see Section 3.3.1).
* This option may reduce the O&M burden on Trading Partners related to the Database Administration for the Staging Table database environment. In this model, EPA would provide the DBA support for the Virtual Node Staging Table components.
* This implementation model will require a secure connection between the Virtual Node infrastructure and the Trading Partner’s infrastructure that will allow for database connectivity. Implementation options for this connectivity are discussed in Section 4. The VN IPT recognizes that this connectivity will require detailed coordination and discussion with appropriate security teams in order to ensure that that connectivity conforms to the Trading Partner’s security policy.
* This implementation pattern mitigates some of the secure connection concerns, but adds significant complexity, cost, and decreases timeliness of data.

### Implementation of Standardized Virtual Node Managed Staging Tables

The implementation described in Section 3.3 is dependent upon a set of staging tables located within the Virtual Node infrastructure. The following bullets detail working assumptions and areas for continued exploration related to the implementation of these Virtual Node Staging tables.

* In order to minimize O&M costs related to implementation of staging tables within the Virtual Node environment, the Virtual Node Staging Table structure for each Exchange Network data flow will be identical across Trading Partners. For example, there will be one standard set of staging table definitions for Facility data that is used for the Facility Identification exchange (FacID). This will allow for the reuse of the XML mapping and/or XML Style sheets across Trading Partners who are using the Virtual Node for the FacID data flow.
* If this implementation model is preferred by Trading Partners on the EN, the teams responsible for VN governance will need to identify what the standard staging table structure will be for any existing data flow that is intended to be migrated to this model within the Virtual Node. If Trading Partners have different staging table structures in existence for a given data flow, this may necessitate an examination and agreement by each VN data flow IPT related to what shared staging table structure to adapt.

## Virtual Node XML Mapping Component

The Virtual Node will implement a XML mapping component that allows for the use of a combination of SQL statements and XML style sheets (XSLT) to allow for direct access to a data source and translation to a resultant XML instance document output that conforms to an XML schema. In addition, this mapping component can be configured for a particular data flow to allow the resultant data set to be provided in other data publishing friendly formats such as JavaScript Object Notation (JSON). Figure 3 depicts the workflow of how the Virtual Node XML Mapping Component connects to a Trading Partner’s data source (either local or Virtual Node staging tables), performs the required mapping and makes the data available in a variety of formats.



Figure 3 Virtual Node XML Mapping Component

# Virtual Node Communication Model

Both implementation models of the Virtual Node platform discussed in Section 3 require a mechanism to allow for interconnecting the Trading Partner’s data source with the Virtual Node environment. The mechanism and configuration of this interconnection will need to be individual agreed to and configured using a combination of resources from both the Trading Partner and EPA’s Virtual Node support team.

The following set of considerations should be accounted for as each Trading Partner seeks to establish this interconnection with the Virtual Node environment.

* **Technical Options for Interconnectivity:** Various methods could be implemented to enable this secure connection between the Trading Partner’s infrastructure and the Virtual Node environment. Security for the VPN connection would be firewall based with traffic only allowed between specific machines (i.e., machine based not user based). Once the connectivity is established, the Virtual Node will be able to make database connections through the secure channel and/or the Trading Partner will be able to make a database connection to the Virtual Node staging table environment. The three preferred methods for establishing this interconnectivity as identified by the VN IPT are the implementation of a Virtual Private Network (VPN) connection, implementation of a Secure Shell (SSH) connection, or use of a cloud brokered connection.
	+ Virtual Private Network (VPN) Connection: A virtual private network (VPN) is the extension of a private network that encompasses links across shared or public networks.[[3]](#footnote-3) This would be accomplished via establishing a virtual point-to-point connection through the use of a dedicated encrypted connection between the Trading Partner and Virtual Node resources.
	+ Secure Shell (SSH) Connection: The Secure Shell Protocol (SSH) is a protocol for secure remote login and other secure network services over an insecure network. The SSH protocol consists of three major components: The Transport Layer Protocol provides server authentication, confidentiality, and integrity with perfect forward secrecy. The User Authentication Protocol authenticates the client to the server. The Connection Protocol multiplexes the encrypted tunnel into several logical channels.[[4]](#footnote-4)
	+ Cloud Brokered Connection: A cloud brokered connection is the connection of disparate network resources via an intermediary external provider. Similar to a VPN or SSH connection, a cloud brokered connection is secured using encryption technologies tunneled through common network protocols such as HTTP. Trading Partners could utilize technologies like this (e.g. Windows Azure Connect) as a potential solution for connections required from their local infrastructure to the VN environment.
* Security agreements and policy considerations: Trading Partners and EPA as the Virtual Node provider will need to collaboratively define these interconnections, rules of behavior related to these interconnections, and establish operational expectations regarding managing these connections over time. These details will be documented in an Interconnection Security Agreement (ISA) as well as a Memorandum of Understanding (MOU) that will be co-authored by both the Trading Partner and the EPA and co-signed by the appropriate signatory authorities within each organization.

# Virtual Node Administration

There are two types Virtual Node Administration that will be performed with the Virtual Node Environment. Much of the basic administration would be done centrally, so that partners are able to concentrate on configuring data flows instead of administering application servers. These two types can be thought as:

* Virtual Node Shared Service Administration: As a shared service created and maintained by the EPA, EPA will provide Virtual Node Administration support that is responsible for the ongoing operations and maintenance of the Virtual Node infrastructure. This will include activities such as:
	+ Web and Application server configuration and patching
	+ Database configuration, tuning and patching
	+ Support of Virtual Node enhancement and/or defect resolution including upgrades to meet new Exchange Network Node specifications as well as upgrades to integrate the Virtual Node to EN services such as ENDS
	+ Performance Monitoring and tuning of the Virtual Node environment
	+ Development, implementation and monitoring of Activity logging services to allow for usage reporting by Trading Partner Node administrators
* Virtual Node Instance Administration: Trading Partners who adopt the Virtual Node will be able to administer and configure various aspects of their Virtual Node instance as needed. This will include activities such as managing metadata about the Virtual Node instance, data flow configuration, and user management.

## Virtual Node Instance Administration

Virtual Node implementers will have the ability to administer various aspects of the Virtual Node instance. This will include the ability to perform the following tasks.

* Administrative Management:
	+ Manage metadata and identifying information regarding the Virtual Node instance that is established for the Trading Partner
	+ Configure and manage data sources used by the Virtual Node
	+ Dashboard and reporting view that displays Virtual Node and dataflow transaction/usage statistics
	+ Features to allow for activity and transaction logging cleanup
* Security:
	+ Manage users and permissions for users related to Virtual Node components
	+ NAAS administration module
* Data Flow:
	+ Data flow configuration wizard to guide Virtual Node administrators and Program office staff through data flow setup and configuration. The wizard will guide Virtual Node administrators through common set up tasks such as data flow name, methods, orchestration patterns, use of data sources, etc. Existing staging tables for plug-ins can be reused to greatly simplify migration of existing flows. SQL Statements can be used to replace entire staging environments for data sources with secure connections using common style sheets and components.
	+ Scheduling interface that allows for scheduling of data flow specific activities
	+ Data publishing capabilities for standard and custom Exchange Network data flows
	+ Ability to manage ENDS Integration per data flow
	+ Ability to orchestrate standard Exchange Network data flow patterns in order to simplify interactions between Trading Partners and applications consuming Exchange Network services
	+ To facilitate data flow changes, the Virtual Node should have a Clone feature that allows implementers to create a new instance of a data flow using the previous configuration of the data flow. The Clone feature should allow for the cloning of the flow within the same development, test and production tier.
	+ A Promotion feature will allow users to promote a data flow configuration between Development, Test and Production environments.

### Virtual Node Administration Portal

In order to perform the administrative tasks described above, Trading Partner Virtual Node Administrators will have access to a Virtual Node Administration Portal (VNAP).

* The VNAP will be a user friendly interface that allows for members of the Trading Partner’s organization to perform the standard set of activities typically required to maintain data flows on the Exchange Network.
* EPA will be responsible for design, development and testing of the VNAP interface.
* Trading Partners will have access to Development, Test and Production environments of the VNAP.
* As changes are made to the VNAP, these changes will be deployed to each environment in successive fashion. Communication and timing of these changes will be in accordance with the MOU and operational agreement established between the EPA and Trading Partners, allowing time for Trading Partners to test, train and address any change management requirements specific to their use of the Virtual Node.

## Virtual Node Security

The following subsections describe various aspects of the overall security model expectations for Trading Partners who implement the Virtual Node.

### Virtual Node Users and Roles

Virtual Node implementers are anticipated to have three types of user roles that will need to be managed for the Virtual Node implementation. All users of the Virtual Node will need a NAAS account.

* **EPA Virtual Node Administrator:** EPA will provide central Virtual Node Administrator support that will assist Trading Partners with Administrative support activities when appropriate. In addition, this role will have permission to implement and make changes to the Virtual Node environments after proper change management and governance procedures have been completed related to the change.
* **Virtual Node Administrators:** Users with the ability to configure and manage global Virtual Node variables (e.g. Virtual Node metadata, data sources, data flow administrator access) as well as configure and manage data flows. Virtual Node Administrators will be able to view and access transaction statistics and history for any data flow on the Trading Partner’s Virtual Node. Sample activities that this role will be able to perform include:
	+ Configure or set up new data sources that can be used by multiple data flows
	+ Setup and administer Virtual Node Data Flow Administrator accounts
	+ Setup and administer NAAS accounts and permissions related to a Trading Partner’s data flows
	+ Node Administrators inherit all permissions granted to Data Flow Administrators
* **Virtual Node Data Flow Administrators:** Users with the ability to configure and manage data flow properties such as data flow services, orchestration, data mapping templates, data flow schedules, etc. Virtual Node Data Flow Administrators will be able to view and access transaction statistics and history for their specific data flow(s) on the Virtual Node. This role allows for Trading Partners to establish a security model where Program Office staff can login and access the Virtual Node in order to manage and/or configure a data flow related to their regulatory program. Sample activities that this role will be able to perform include:
	+ Modify SQL statements/mapping
	+ Configure the virtual mapping template or style sheet
	+ Trigger a flow that is not on a set schedule
	+ View historical transactions for that data flow
* **Trading Partner NAAS accounts:** Users who have access to invoke a Trading Partner’s data flows. Virtual Node Administrators will have the ability to create new NAAS accounts and assign data flow permissions to these NAAS accounts through the Virtual Node interface.

### Virtual Node Security Module

The Virtual Node will include a Security Module that will allow Virtual Node implementers to perform the following tasks:

* View and manage Virtual Node Administrator accounts
* View and manage Virtual Node Data Flow Administrator accounts
* View and manage NAAS accounts and associated permissions related to a Trading Partner’s data flows

# Implementation and Operational Recommendations

This section discusses various implementation and other operation recommendations identified by the Virtual Node IPT members including:

* Approaches for transitioning to the Virtual Node
* Resource Planning for Virtual Node implementers
* Help Desk requirements
* Training and documentation requirements
* Change Management and Governance

## Transitioning to the Virtual Node

The Virtual Node IPT examined the implementation of an Exchange Network Virtual Node from the perspective of how new data flows could be implemented in the platform as well as how existing data flows implemented by a Trading Partner could be migrated to the platform. This section of the Virtual Node Guidance and Recommendation document discusses the approaches that Exchange Network Trading Partners will need to examine related to transitioning existing data flows the Virtual Node platform.

The following transition planning recommendations have been identified by the Virtual Node IPT:

* Virtual Node implementers should leverage existing investment by reusing existing plug-in components as much as possible. For most Trading Partners on the network this will primarily be focused on reuse of an existing mapping template, XML style sheets and staging tables. Either implementation pattern discussed in Section 3 of this document can be utilized in a manner that will allow for this reuse to occur. Note that in the event that a Trading Partner wants to leverage a Virtual Node staging table method, coordination at a Data Flow IPT level may be necessary in order to identify what the common staging table model will be for all Trading Partners who are likely to use the same model.
* Virtual Node implementers should anticipate the following key milestones related to transitioning to the Virtual Node for a given data flow. Note that some of these activities will only be required during the initial setup of the Virtual Node for the Trading Partner.
	+ Setup of Virtual node instance through coordination with EPA
	+ Decision on interconnectivity model through coordinated discussions with Trading Partner and EPA security and technical teams
	+ Drafting and execution of ISA/MOU materials
	+ Implementation and testing of interconnectivity model
	+ Setup of initial Virtual Node Administrator accounts
	+ Training of Virtual Node Administrator
	+ Training of Virtual Node Data Flow Administrator (if applicable)
	+ Configuration of data flow within Virtual Node Administration Portal including any adjustments to ETL or data mappings required based on the implementation approach decided upon by the Trading Partner and data flow IPT.
	+ For new data flows, Trading Partners can use SQL Statements rather than staging environments for data sources with secure connections using common style sheets. For data flows with staging table components, Trading Partners may need to map source system data structures to a staging table and then implement SQL statements for extraction from those staging tables for the data flow that is being implemented.
	+ Testing and deployment activities related to the data flow configuration through three tiered Virtual Node architecture
	+ Resource planning for ongoing support of the data flow configuration
	+ Communication with existing Trading Partners who access the previous implementation of the data flow in order to appropriately direct them to the new endpoint for the Virtual Node implementation of the data flow (if applicable)
* Trading Partners may opt to continue to support a local node infrastructure for certain data flows and/or during a transitional period.

## Resource Planning

The Virtual Node IPT recognizes that the adoption of the Virtual Node will represent a business model change for Trading Partners currently involved on the Exchange Network. As such, it is important to recognize the impacts that this will have on existing resources for Exchange Network partners. In addition, as the Virtual Node is intended to reduce the barrier to entry for new Trading Partners on the EN, it is important to set accurate resource requirement expectations with those organizations related to use of the Virtual Node.

The following resource planning recommendations have been identified by the Virtual Node IPT:

* **Virtual Node User Base:** Trading Partners should expect to have two types of Virtual Node users who interact with the Virtual Node on a regular basis (the Virtual Node Administrator and the Virtual Node Data Flow administrator). These two roles could be performed by a single individual. However, as a result of the fact that the majority of the O&M work associated with a Trading Partner’s core node presence will be managed by EPA, it is possible that the majority of use by the Trading Partner may be by team member(s) who are more at the Program Office level responsible for a specific data flow. As such, IPT members noted that the Virtual Node will allow for a closer “connection” between Program Office staff and their data flow than historical models where a central IT organization was more intimately involved in the general setup, maintenance and troubleshooting of data flows.
* **Setup Resources:** During initial setup and configuration of a Trading Partner’s Node presence, the IT organization of the Trading Partner should expect to be involved in the provisioning of global node variables as well as working through the security model for interconnectivity. It is expected that the implementation and execution of associated ISA/MOU documentation for this interconnectivity will require coordination between IT, security and Trading Partner management teams.
* **Transition Resources:** If a Trading Partner is planning on migrating existing data flows to the Virtual Node platform, a specific set of resources will be required to support the activities associated with this transition. Section 6.1 of this document describes the typical transition activities that the Virtual Node implementer should plan on supporting. Example resource expectations to anticipate during these activities include Node Administrator support for user configuration, technical Program Office support to address any data mapping changes, Program Office business user support to test and validate results, outreach/documentation support for any integration points with other Trading Partners, and possible coordination with data exchange partner testing teams (e.g. testing the results of a migrated national system flow with an EPA Program Office team).
* **Training Resources:** Trading partners should expect upfront resource requirements related to training a resource(s) on Virtual Node Administration tasks. As new data flows are implemented on the Virtual Node that involve new business units within a Trading Partner’s organization, the Trading Partner should plan for a small upfront investment in the training of this resource (training and documentation resources that are expected to be available are noted in Section 6.4).
* **Support Resources:** Ongoingsupport for the core Virtual Node shared infrastructure will be provided by EPA. Virtual Node implementers should plan on ongoing support activities related to user management requests, supporting interconnectivity with the Virtual Node environment, and data flow support. The number of resources required to perform these tasks will vary based on the number and types of data flows implemented by the Trading Partner.
* **Change Management Resources:** It is expected that changes will occur to the Virtual Node platform and associated data flows over time.
	+ **Data flow changes:** Resource requirements to manage data flow changes are expected to be similar to existing models except for the fact that any changes related to the core node implementation will be managed by EPA. Implementation of data flow mapping changes and associated testing will remain the responsibility of the Trading Partner.
	+ **Virtual Node Platform changes:** Virtual Node implementers will have the opportunity to assess and manage the impact of Virtual Node platform changes (e.g. User interface, node specifications) as they are deployed across the Virtual Node development, test and production environments. Virtual Node implementers should expect the need for occasional change management activities to be required as a result of adopting this shared service model. This will include coordination, documentation review and the possible need for small targeted training or outreach to understand the changes being implemented.
* **Data Flow Configuration Support:** IPT members noted the need for access to contractor resources that would have the necessary skill sets required to configure data flows in certain cases where internal staff members with these skill sets are not available. EPA is currently exploring options that can be made available to Trading Partners to get access to these types of resources either through the support provided via the Exchange Network Help Desk or other alternative models.

## Help Desk Requirements

The Virtual Node IPT recognizes that the adoption of the Virtual Node will require a centralized Help Desk model in order to support Virtual Node implementers during initial setup, transition, creation of new data flows and ongoing O&M support activities.

The following help desk recommendations have been identified by the Virtual Node IPT:

* The existing Exchange Network Help Desk provides an existing model that can be naturally extended to provide Virtual Node support to implementers.
* Technical support for data mapping will likely be maintained by staff within the Trading Partner’s organization.
* EPA and the Exchange Network should consider a “for-hire” service that states and tribes could contract with for more advanced help in the event that Trading Partner staff members require deeper support for data flow configuration beyond basic support that would be expected from the standard Exchange Network Help Desk model. It is possible that this could be a further extension of the existing Exchange Network Help Desk model that allows Trading Partner business Subject Matter Experts to be paired with “for-hire” Virtual Node technical support teams to implement data exchanges.
* The Exchange Network Help Desk should be available to states and tribes that need support for national system flow set up (perhaps via an EPA/ECOS agreement but that will need additional discussions). IPT members recommend establishing standard connections between the resources on the Exchange Network Help Desk and resources already available via the EPA program offices to facilitate transition.
* A bulk of the hours set aside for the help desk will be for states and tribes to implement existing flows.

## Training and Documentation

The Virtual Node IPT recognizes that the adoption of the Virtual Node platform will require upfront and ongoing training as well as detailed documentation in order to manage successful adoption of the platform.

The following training and documentation recommendations have been identified by the Virtual Node IPT:

* During the initial launch of the Virtual Node platform (and on a periodic bases thereafter), EPA should offer webinars that target key training topics such as setup, data flow configuration and data flow O&M activities. Webinars should result in recorded material that can be made available to the Virtual Node user base on demand.
* Training materials and documentation should be made available online either on or accessible from the Exchange Network website. Where and if possible, material related to online training in a modular format, should be published by EPA.
* Detailed documentation related to the use of the Virtual Node should be created and kept up to date. This documentation should be made available on the Exchange Network website.
* Release notes should be provided as any changes to the Virtual Node are made.
* A Knowledge Transfer/discussion forum should be established on Exchange Network website resources to enable collaboration between Virtual Node users.
* If possible, IPT members were interested in establishing a feature within the Virtual Node that would allow Trading Partners to share sample data flow configurations. This needs to be explored further but ideas such as the ability to share data flow templates, grant access to development environments or the creation of a shared sandbox environment were possible suggestions.

## Change Management and Governance Considerations

The Virtual Node IPT recommendations related to Governance are primarily focused on how to manage and coordinate change within the Virtual Node Platform user base. Two primary principles related to this governance were agreed to by the Virtual Node IPT during the team’s formation:

* Existing EN Governance will be sufficient to address issues extending beyond the Virtual Node implementation and features. For example, changes related to the creation of REST specifications, or changes to data flows are managed within Network Technology Group or data flow IPT forums and should continue to be managed in that fashion. The data flow configuration and implementation in the Virtual Node will follow these team’s recommendations.
* Tribal representation related to implementation, use and adoption of the Virtual Node is a key component to its success.

The following change management recommendations have been identified by the Virtual Node IPT:

* The IPT recommends considering change management from the perspective of three types of changes. Recommendations related to how to manage these changes for Virtual Node implementers are included as sub bullets under each item.
	+ **Data flow changes:** These include changes to data flow specifications such as the adoption of a new XML schema for a data flow.
		- Trading Partners who use the Virtual Node will implement changes to a data flow when they are ready to adopt the changes. These will not be automatically pushed to each Virtual Node instance.
		- Each version of a data flow can be a separate instance of that data flow for the Virtual Node implementer. As a result, Trading Partners should be able to configure a new version of the data flow while maintaining the previous version.
		- To facilitate data flow changes, the Virtual Node should have Clone and Promote features that allow implementers to create a new instance of a data flow using the previous configuration of the data flow. The Clone feature should allow for the cloning of the flow within the same development, test and production tier. In order to promote a data from development to test to production, Virtual Node users should use a separate Promote feature that will replicate the data flow configuration and address and data source changes required.
		- The retirement or deprecation of data flow versions will be handled through the existing EN governance process and is outside of the purview of the Virtual Node IPT.
	+ **Virtual Node feature changes:** Virtual Node feature level changes include items such as enhancements or the addition of new functionality to the Virtual Node Administration Portal (VNAP).
		- Existing Exchange Network governance bodies (e.g. Network Technology Group) can be used to solicit and collect user input and joint decision-making on feature changes for the Virtual Node. This will ensure that users have an open forum for discussion and a stake in the management of the VN as a shared central service.
		- Virtual Node implementers will receive prior notification of upcoming releases and changes related to the VNAP.
		- Virtual Node implementers will use the time afforded during the deployment of these changes to the Virtual Node development and testing regions to train impacted users.
		- EPA will provide updated documentation that can be used to facilitate this training.
		- If the impact is substantial, EPA will provide webinar based training.
	+ **Primitive/Service Level Changes:** Changes to the Node Protocol and Specifications are considered Primitive/Service level changes (e.g. changes to service specifications such as Query or Execute).
		- These services would only change when the Protocol or Specifications change according to existing mechanisms within the EN (e.g., through the ENLC).
		- The EPA Virtual Node support team should consider utilizing capabilities of the EPA cloud environment to facilitate the implementation of new primitive level changes to the node specifications.
* Timelines for notification to Virtual Node implementers will be documented and agreed to in the template MOU that will be cosigned by EPA and the Virtual Node implementer.
* Changes to the Virtual Node will be implemented using industry standard best practices across development, test and production environments, allowing for assessment of changes, regression testing and appropriate quality assurance activities.

## Memorandum of Understanding Recommendations

The Virtual Node IPT recognizes that need for the appropriate mechanisms to be adopted to set expectations between EPA and each organization that adopts the Virtual Node. As a result the following recommendations are put forth by the IPT related to content that should be agreed to within a Virtual Node Memorandum of Understanding.

* Some MOU content will be agreed to at the Exchange Network level and will be adopted “as-is” by each Trading Partner. Content of this type will not be subject to change in each adopted instance of the MOU. For example, appropriate notification timelines for changes to the Virtual Node Administration Portal will be standardized across all MOUs.
* General content within the MOU will include the following topics:
	+ Description of VN and its purpose
	+ Clarity on who the agreement is between
	+ Funding responsibilities
	+ Hosting location
* Operational support content within the MOU will include the following topics:
	+ Expectations for who provides what support
	+ Level of service/up-time requirements
	+ Change control policy and schedule for communication
	+ Backup and restore policies
* Security expectations and associated rules of behavior for each party should be documented in the MOU
* A standard dispute resolution process will be documented in the MOU
1. <http://www.exchangenetwork.net/virtual-node-ipt/> [↑](#footnote-ref-1)
2. Refer to Section 6.5 for recommendations related to managing changes to the Virtual Node related to Node or data flow specification changes. [↑](#footnote-ref-2)
3. Microsoft Technet. "Virtual Private Networking: An Overview" (<http://technet.microsoft.com/en-us/library/bb742566.aspx> ) [↑](#footnote-ref-3)
4. Network Working Group of the IETF, Secure Shell Protocol Architecture (<http://tools.ietf.org/html/rfc4251>) [↑](#footnote-ref-4)