Conceptual Solution Architecture

Data Management Tools - Master Data Management (MDM) Platform

|  |  |
| --- | --- |
| INVP # (if known) | 6252US |
| ITLT Sponsor | Narayan Devireddy and Emer McNally |
| IT Business Partner | Ed Brodsky |
| Domain Architect | N/A |
| Information Architect | Nishit Ajwaliya |
| Solution Architect | Sean McAuley |
| Infrastructure Architect | Ramya Dama |
| Security Architecture | Reza Maserrat , Igor Barshteyn |
|  |  |

**Document Control**

Classification INTERNAL USE / CONFIDENTIAL

Owner Nishit Ajwaliya

Author Nishit Ajwaliya

Creation Date: August 10, 2021

Version 0.1

Status DRAFT

Location Architecture Group Site

Template version v2.12

**Contents:**

1 Business Vision 3

2 Business Scenarios 4

2.1 Use Cases 4

3 Stakeholders and their Concerns 4

4 Constraints 5

5 Business Architecture 6

6 Users and User Experience Expectations 7

7 User Impact/Expectations 7

8 Current State 8

9 Assumptions 8

10 High Level Solution Architecture 8

11 Technology Disposition Implications 10

12 Integration Architecture 10

13 Information Architecture 11

13.1 Data Subject areas 11

14 End User Compute Architecture 11

15 Infrastructure /Hosting/Network Architecture 13

16 Performance Requirements 13

17 Backup Requirements 13

18 Disaster Recovery Requirements 14

19 Security Architecture 15

19.1 Information Classification / Criticality 15

19.2 Regulatory Impacts 15

19.3 Baseline Security Requirements (BSR’s) 16

19.4 Security Pattern 16

19.5 Security Tower Engagement 16

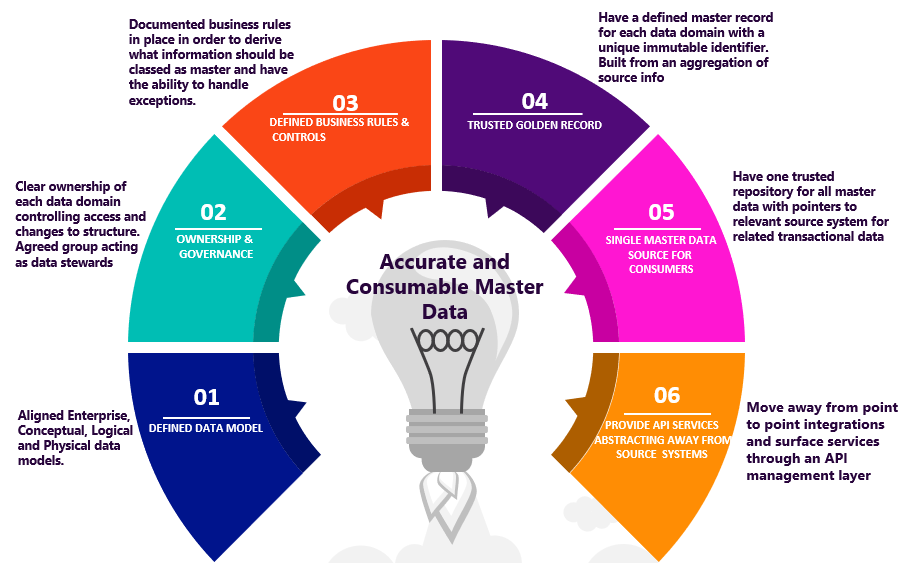
19.6 Additional Security Patterns 16

20 IT Commercial 16

# Business Vision

The current state of Data at National Grid poses significant challenges for the company. It exposes us to security risks, has already impacted our major programs, and necessitates costly manual efforts to deliver our daily operational services. “Fixing Data” holistically (and no longer as one-off efforts) is imperative and urgent. This paper seeks to build on our agreed actions to hire a Chief Data Officer, continue to improve our Data Governance & Stewardship building off the progress of the Data BMS, and proposes additional steps to “Fix Data.” We would appreciate your feedback and commitment to arrive at a clear, unified, and decisive path forward.

The business vision is to leverage a single, trusted view of master data as shown below:



Reltio Master Data Management is a selected platform solution for the National Grid enterprise Master Data Management (MDM) to manage consolidated master data. The key benefits of Reltio MDM platform is following:

* Centralized multi domain master data repository
* An enterprise unique persistent identifier for each master data domain entity
* Avoid master data duplication
* Increase master data accuracy
* Improve master data quality
* Improve master data consistency and data standardization
* Provide better data compliance
* Enable data governance and data stewardship
* Offer better secure data access
* Provide an API service for consumers of this curated data

Note: The project scope is focused on the technical implementation of standing up the Master Data Management platform.

# Business Scenarios

The project scope is focused on the technical implementation of standing up the Master Data Management platform and does not have business scenarios. As use cases get onboarded into the tools being implemented, business scenarios will be populated.

The overall focus of the engagement is to establish a Master Data Management (MDM) for future business scenarios and tools that will allow maintaining integrated Master Data in one platform that will be used by the various business.

## Use Cases

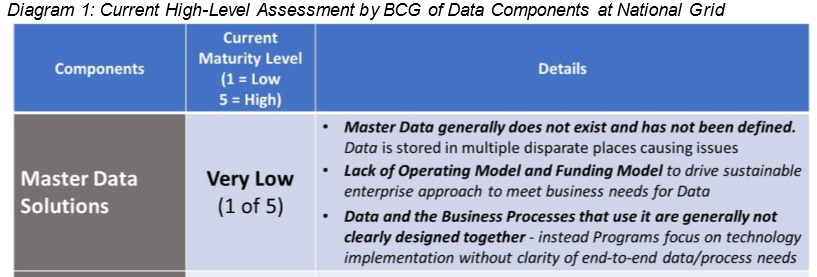
The primary scope of the project is to focus on the technical implementation of the Master Data Management (MDM) platform. When the specific master data domain is implemented, following use cases will be covered:

|  |  |
| --- | --- |
| **Use Case** | **Benefit** |
| MDM1 – Centralized data repository for multi-domain master data | All master data from multiple domains in one place, one governance and shared enterprise control process |
| MDM2 – System of record for unique enterprise persistent identifier for each master data entity | MDM can create a single unique identifier for each master data entity like Worker, Position, Customer, Asset, and Location etc. that is required by multiple business processes |
| MDM3 – Allow data quality improvements, remove data duplication, and allow data steward to see data issues and help them correcting them | Consistently maintain data quality and remove duplicate records of the Worker, Position, Customer, Asset, and Location master data entities |
| MDM4 – Create and maintain golden record of master data and sync cleanse and standardize data back to source systems of record | Remove data discrepancy between different source systems and have an authoritative and a reliable source of Customer, Workforce, Asset master data that solves the one biggest pain point of multiple business areas |
| MDM5 – Source for Master data for Common Enterprise API/Services, MuleSoft, CDP/Data Lake (Snowflake), or any other downstream systems and improve performance by reducing dependency of source systems | Consistent clean accurate master data for Common Enterprise API/Services, MuleSoft, Cloud Data Platform/Data Lake (Snowflake), or any other downstream systems |

Note: The project scope is focused on the technical implementation of standing up the Master Data Management platform.

# Stakeholders and their Concerns

As per the “Fixing Data” paper presented in April 2020 to executives by Chief Data Officer Charles Zentay at a time, unless we “Fix Data” our security, major programs, and financial and operational performance are at risk. The paper was based work by Boston Consulting Group BCG). As per the paper National Grid rates “low” or “very low” across the three major components of Data (Master Data Solutions, Data Governance, and Data Tools & Platforms). BCG’s study, that looked into a subset of the Data domains at National Grid, estimated the costs of bad Data are $73-98M. Given the limited scope of their review, they concluded that the full cost is likely much higher.”



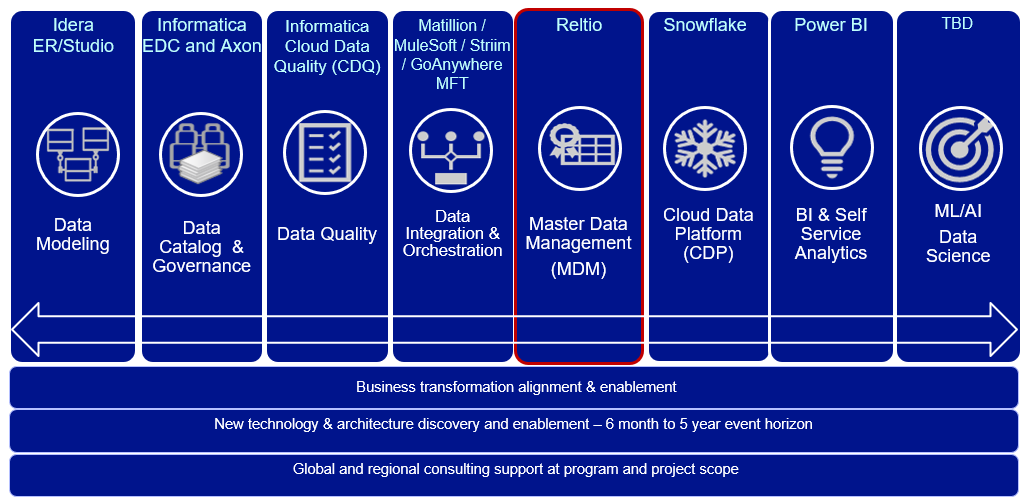
To addresses these issues, we need three key components to “Fix Data”:

1. **Master Data Solutions – for each data domain, establish a data master**
2. Data Governance & Stewardship – Business ownership of data, accelerated by BMS
3. Data Tools & Capabilities – One, common approach to Enterprise Data solutions

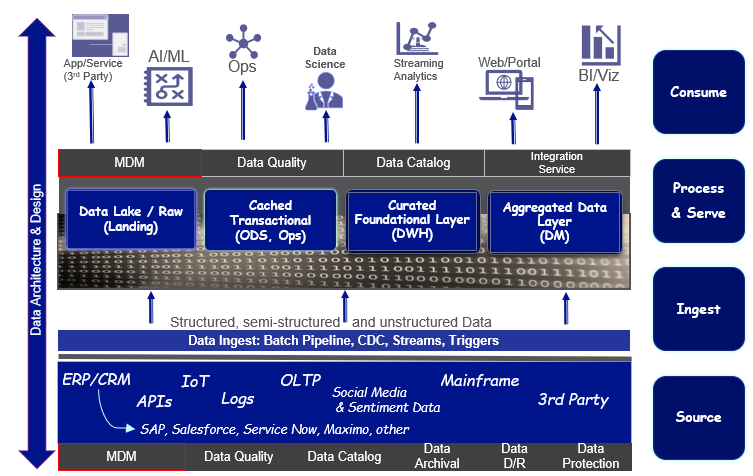
# Constraints

* Some source systems might be resource and/or technical constrained to support the added feed and to build new real-time or near real-time interfaces for the MDM.
* Some of the source systems are changing with the implementation of various projects/programs. Will have to carefully plan the build of data ingestion interfaces from these systems to minimize the impact and rework.

# Business Architecture

Master Data Management is one of the key business data capability pillars in the Enterprise Data Platform as shown below: 

The pillars outlined above cooperate in a secure manner aligned with National Grid’s policies on identity management and role-based access control, so that all levels of regulatory compliance are met, at every stage of the development, testing, deployment, and operational lifecycle.



# Users and User Experience Expectations

In Scope User Personas

|  |  |  |
| --- | --- | --- |
| Persona / Users | Interacting Internal/External  to Corp Network | User Impact or User Expectations |
| Regulatory | External | * Improved speed of delivery for business functionality * Master data quality checks and validations implemented * Higher quality master data available for making business decisions * Master data available for use in predictive analytics * Improved transparency and responsiveness. * Improved data management capability and broader visibility for US Customer, Global Workforce, US Electric Business Unit (EBU), US Gas Business Unit (GBU), UK Electric Transmission etc. * Visibility to master data from several systems in a single interface helps business make informed decisions and better manage operational processes |
| Corporate Finance | Internal/External |
| Security | Internal |
| Field Workforce Management | Internal/External |
| Control Room Operations – Distribution and Transmission | Internal |
| Asset Management | Internal/External |
| Distribution Operations | Internal/External |
| Distribution Planning and Asset Management (DPAM) | Internal/External |
| Customer Service | Internal |
| Customer | External |
| Emergency Planning Group | Internal/External |
| Wholesale Electric & Environment | Internal/External |
| Interconnection Group | Internal |
| Energy Procurement Group | Internal/External |

# User Impact/Expectations

The business demands for the Master Data Management are as follow:

***US Business Domain Demand​***

* US Customer – MDM requirement to build high quality, consistent, accurate, and complete master customer and meter data from numerous US customer systems of record like CRIS, CIS, CRM etc. and make them available to enterprise e.g. Customer Transformation, GBE, AMI, Digital Enablement (On My Way), Energy Efficiency/Clean Energy initiative, Smart Target initiative, Consolidation of Customer Data Sources to scalable centralized repository etc.
* US Asset – MDM requirement to build high quality, consistent, accurate, and complete master asset data from numerous US asset systems of record like Power Plan, GIS, Cascade, Computapole, STORMS, Maximo etc. and make them available to enterprise eg AMI, GridMod, Digital Enablement (FutureNow, VMO) etc.

***UK Business Domain Demand​***

* UK Customer – MDM requirement to build high quality, consistent, accurate, and complete master customer data from numerous UK B2B customer systems of record CRM, SAP etc. and make them available enterprise wide eg ESO etc.
* UK Asset – MDM requirement to build high quality, consistent, accurate, and complete master asset data from numerous UK asset systems of record like Ellipse, FieldReach, GeoGrid, ET Power Factory etc. and make them available to enterprise eg ESO, RIIO T2 based UK GT and UK ET initiatives etc.

***Global Business Domain Demand***

* Workforce – MDM requirement to build high quality, consistent, accurate, and complete master workforce (Employee, Contingent worker/Contractor, and Managed Service Provider/MSP workers) data from numerous workforce systems of record like MyHub/ SuccessFactor, FieldGlass, US/UK SAP etc. and make them available to enterprise e.g. Workforce Data Domain, IAM, GBE, Digital Enablement (On My Way) etc.

The project scope is focused on the technical implementation of standing up the Master Data Management platform.

# Current State

Master data management is kind of a new data capability for the NationalGrid so there is not much to report on current state.

UK Finance is using very limited Master Data Management functionality of SAP Master Data Governance tool which will be deprecated through introduction of this new standard of Reltio MDM.

# Assumptions

|  |  |
| --- | --- |
| # | Assumption |
|  | Reltio MDM contract includes separate US and UK instance but we stand up only US SaaS instance right now |
|  | Separate document will be created for MDM implementation of any business domain by each domain specific initiative |
|  | Reltio currently supports only AWS and GCP based SaaS solution so we will be using AWS cloud solution until they start supporting Azure next year (2022) |
|  |  |
|  |  |

# High Level Solution Architecture

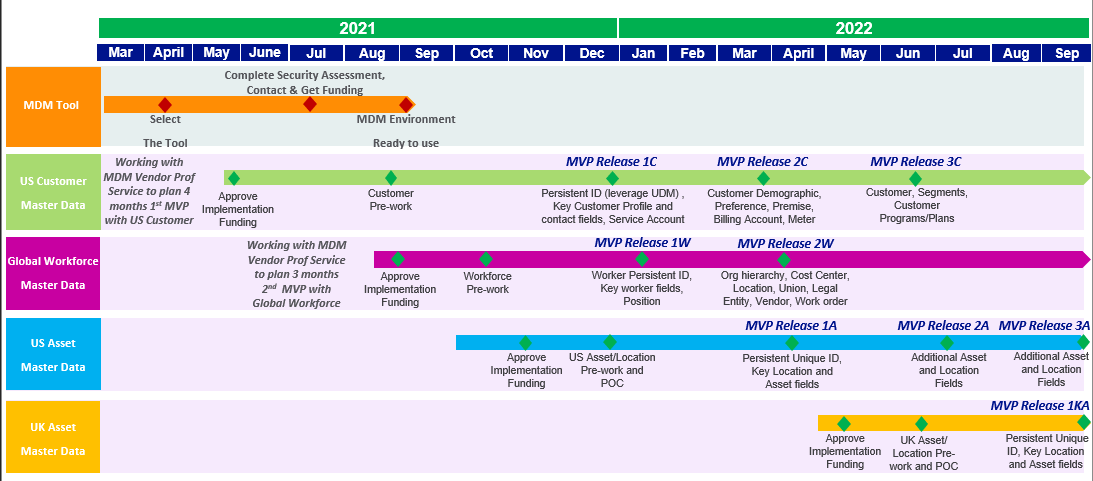


Solution Approach

* Requirements & Design to stand up the Reltio MDM platform
* Reltio to stand up SaaS Development, Test and Production SaaS MDM tenants
* Establishing Enterprise Level NG connectivity and Single Sign On (SSO) with Development, Test and Production Reltio SaaS MDM Environments
* Complete basic minimum MDM configuration
* Implement security recommendations
* Security testing
* Integration testing
* Disaster recovery plans

**Master Data Management Proposed Roadmap**

Please see the proposed draft Master Data Management roadmap below once the MDM platform is available to use:



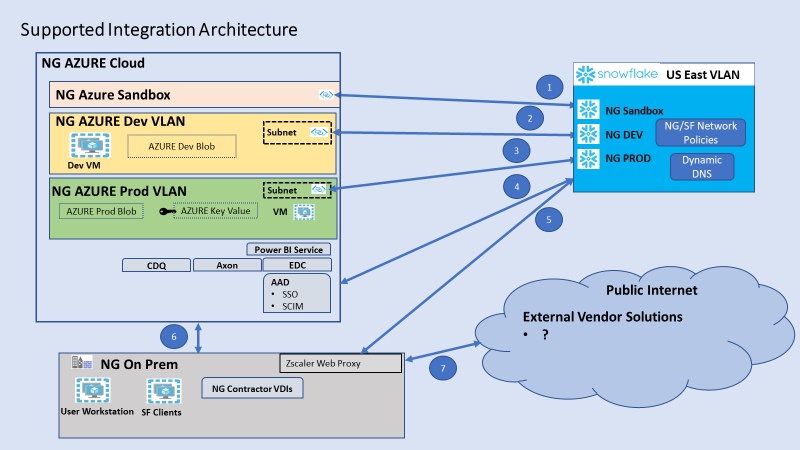
Note: The above roadmap covers the standing up MDM environment as well as proposed timeline to implement first few data domains MVPs. The project scope is focused on the technical implementation of standing up the Master Data Management platform.

# Technology Disposition Implications

|  |  |  |
| --- | --- | --- |
| Technology | Capabilities | Rationale |
| Reltio MDM | Master Data Management | New enterprise strategic master data management platform |

# Integration Architecture

As a data warehouse software as a service, Snowflake at National Grid must integrate with many types of other software services or platform. The diagram below



1. Azure private link connecting NG Snowflake Sandbox account to the NG Azure Sandbox
2. Azure private link connecting NG Snowflake Dev account to the NG Azure Dev VLAN
3. Azure private link connecting NG Snowflake Prod account to the NG Azure Prod VLAN
4. Azure to Azure link flowing traffic between NG owned Azure services and hosted applications to the Snowflake AZURE VLAN using public IP addresses
5. Connection from Zscaler across the public Internet to the Snowflake Azure VLAN
6. Connection between NG On-Prem to Azure
7. Connections between public internet clients and Zscaler

National Grid has three Snowflake accounts. Each Snowflake account has a public and private URL. NG will support access through Snowflake either through a private link, through the NG proxy server, or using public URLs to connect AZURE to AZURE.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Account | | Public URL | | Private URL | NG ENDPOINT | |
| Production | tbd | | tbd | | |  |
| Development | tbd | | tbd | | |  |
| Test | tbd | | tbd | | |  |

Snowflake supports creating network policies at the Snowflake account level. A policy can whitelist and blacklist IP addresses. Inside of an account, there can be a default policy for the account and user specific policies.

National Grid supports user authentication through their Azure Active Directory (AAD). A few user accounts will have access outside of SSO as part of DR plans to provide Snowflake access in the event that AAD is not available.

Snowflake will have access to blob stores on NG Azure subscriptions. Currently there are no plans to access blob stores outside of the NG network.

The network link between the NG on-prem network and Azure Cloud has limited capacity. All large data flows can occur only on the cloud. This is a physcial limitation of the link and not of the used software.

# Information Architecture

The project scope is focused on the technical implementation of standing up the Master Data Management platform and does not have any impact to Information Architecture. When a specific data domain is implemented, specific information architecture will be provided.



The operational source systems support the actual business processes and store the Master Data as well as the transaction data.

The Master Data Management platform is at the centre of managing master data. MDM platform gets master data from multiple operational source systems. It performs basic profiling on them, standardise them, cleanse them, enrich them with third party data, improved the data quality as per business rules, match and merge them, create an enterprise persistent unique identifier, and finally a golden master record is created. These cleanse master data can be sync with operational source systems.

The cloud data platform is one of the main final recipients of the Master Data, generated and managed by the supporting components. The cloud data platform is the enterprise consolidated repository of both master and transactional data for the company. The Master data is synced from the MDM platform in near-real time.

The business intelligence reporting is performed using data from the cloud data platform data. The quality of the reporting generated is fully dependent on the quality of the Master Data and transaction data into the cloud data platform.

The third-party vendor data like D&B, industry standard reference data and other are used to enrich the master data.

The downstream applications/systems can get the quality master data in real-time via API or in batch from MDM repository.

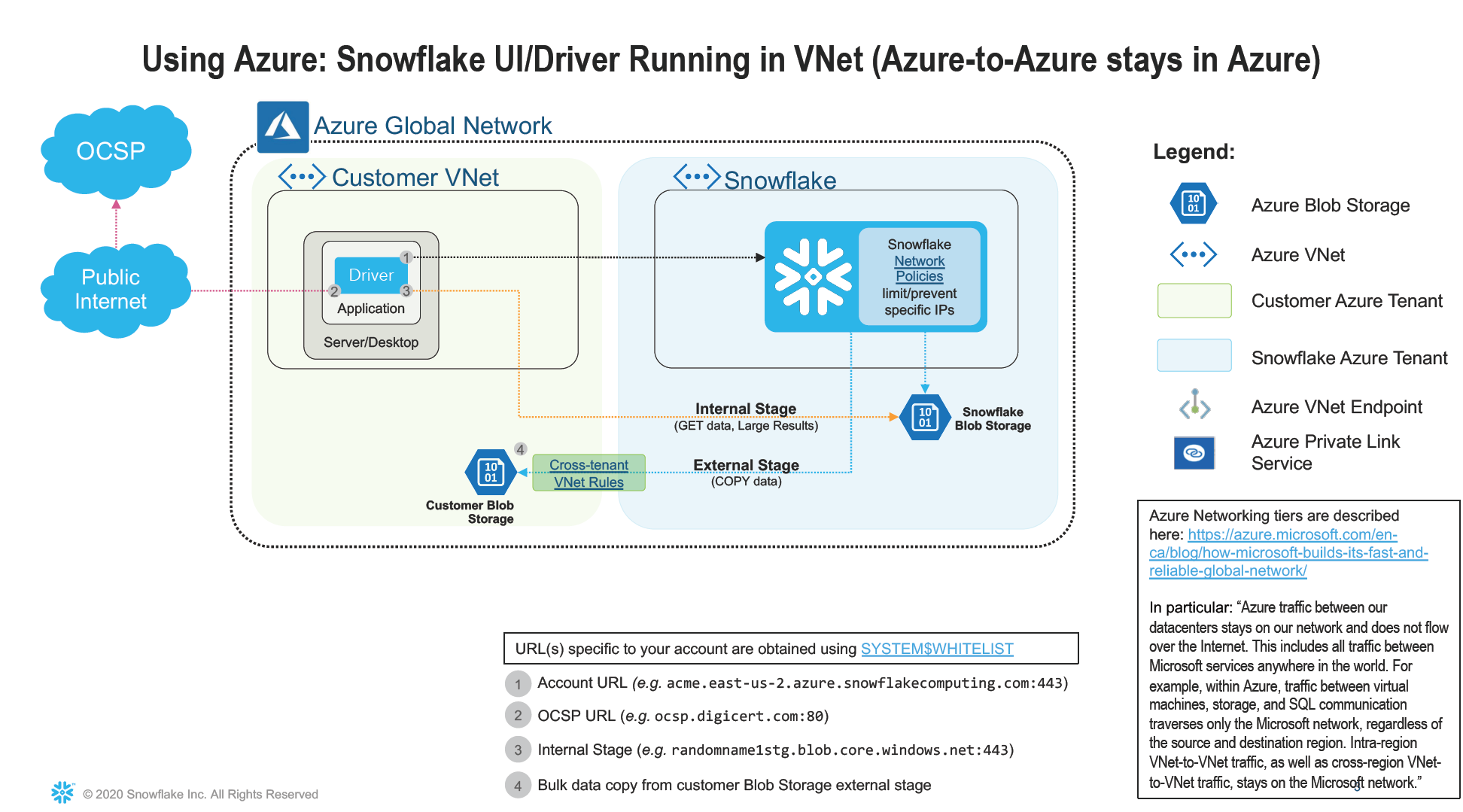
## Data Subject areas

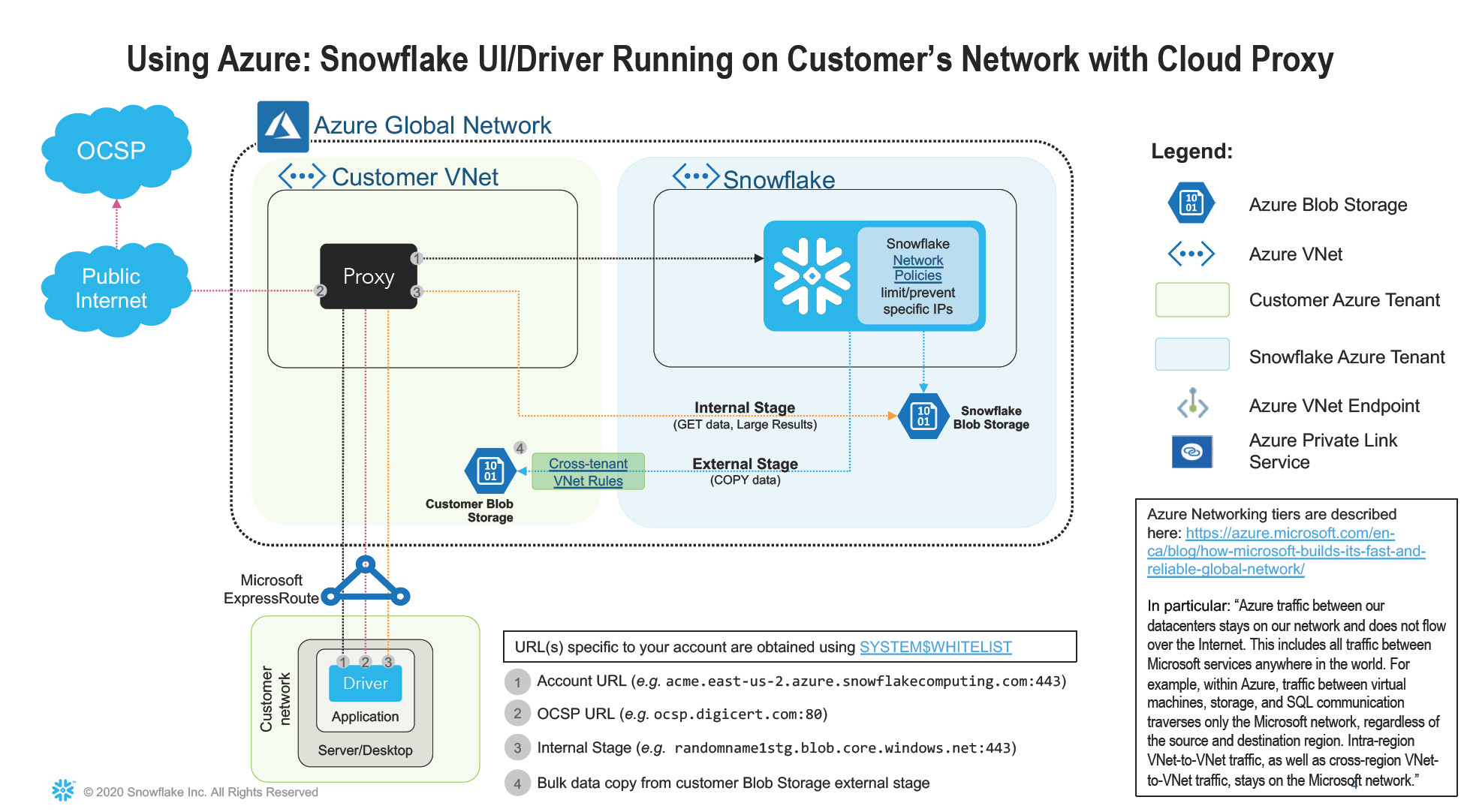
The project scope is focused on the technical implementation of standing up the Master Data Management platform so data subject areas are not in scope for this project as no data will be imported.

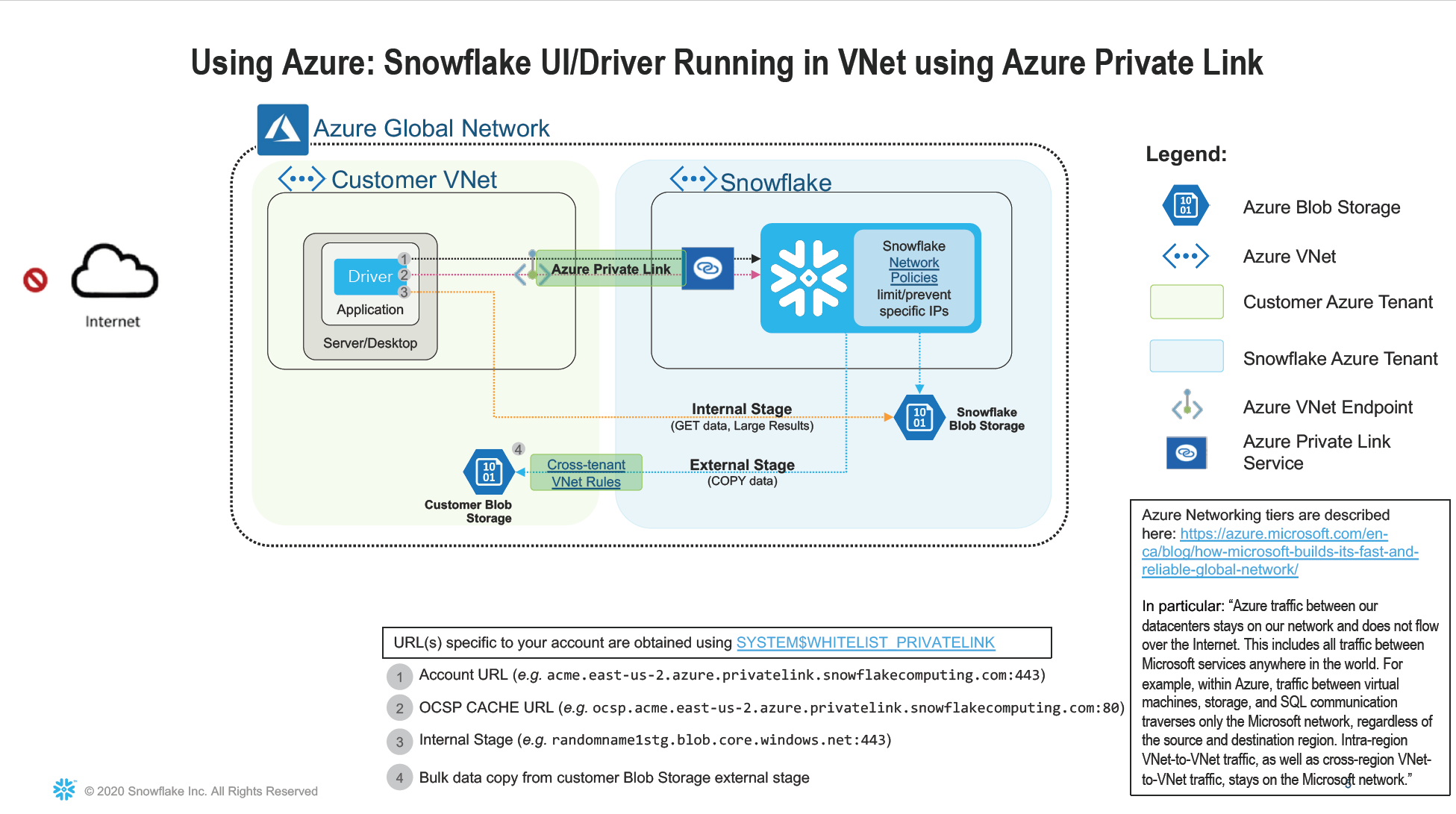
# End User Compute Architecture

Business users access Snowflake Cloud DB from VDI/laptop on VPN using SQL client to run queries, analyse data and generate KPIs.

The following pictures show the client architecture for Azure to Azure connections. (Source Snowflake)







Business users access Informatica Tools from VDI/laptop on VPN using web browser to Informatica hosted web servers..

# Infrastructure /Hosting/Network Architecture

The architecture diagram under section 12. Integration Architecture applies here.

# Performance Requirements

No Specific performance requirements given.

* Authorized users should be able to access Master Data Management platform via Reltio web interface using web browser
* Authorized users can access data via Reltio API

# Backup Requirements

With Snowflake, the subject of backup requirements is a component of disaster recovery requirements. Snowflake provides data redundancy as a core part of their service.

Snowflake addresses data redundancy by:

* Snowflake uses unlimited “blob” storage for database and file stage storage, which provides “eleven nines” (99.999999999) of durability. All storage is stored redundantly in multiple availability zones (AZ).
* Snowflake supports the ability to query as of a specific time in the past. The default is to keep all data for one day. Snowflake calls this feature “time travel”. The National Grid version of Snowflake permits setting time travel for up to 90 days at the account, database, schema, or table level. Each application hosted on Snowflake must set their own time travel requirements.
* Snowflake has a concept they call “fail-safe”. Data stored on Snowflake is retrievable in a disaster for up to 7 days. In order to retrieve fail-safe data, National Grid needs to open a customer support ticket with Snowflake. It can take several days to recover this data. Applications hosted by National Grid on Snowflake should not design their disaster recovery strategies using fail safe.
* Snowflake provides the ability to replicate across Snowflake Regions. This capability is supported across multiple cloud providers. At this point, National Grid is not setting replication.

# Disaster Recovery Requirements

There are two primary objectives that pertain to disaster recovery and business continuity.

* RTO – Recovery Time Objective is the amount of time that an organization permits before bringing a service up and fully available.
* RPO – Recovery Point Objective pertains to the amount of data and functionality that must be operational when recovering an application.

From a platform perspective, the goal is to keep the platform up and running at 99.95% as provided by Reltio.

The goal is to recover all data in the event of a disaster. Uncommitted database transaction shall fail and need to be resubmitted. It is the responsibility of application team to address the integrity of business transactions.

# Security Architecture

## Information Classification / Criticality

*Refer to ISMS129 Cloud Security, ISMS 102 Information Classification*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Public | Internal Use | Confidential | Strictly Confidential |
| Operationally Critical | Forbidden | Private | Private | Private |
| Critical | Private / Hybrid | Private / Hybrid | Private / Hybrid | Private |
| Core | Public / Private / Hybrid | Public / Private / Hybrid | Public / Private / Hybrid | Public / Private / Hybrid |
| Efficiency & Performance | Public / Private / Hybrid | Public / Private / Hybrid | Public / Private / Hybrid | Public / Private / Hybrid |

Information maintained by the EDP is restricted to internal use. The platform shall host applications that host strictly confidential data. Data stored in Snowflake is encrypted. Data stored on Azure blob stores is not under control by the Snowflake platform. Blob store access credentials shall be stored in Snowflake using Snowflake objects called storage integrations. Access and use of the storage integrations is controlled by Snowflake security roles. The credentials are not accessible by end users.

## Regulatory Impacts

Data imports are out of scope for this engagement.

|  |  |  |
| --- | --- | --- |
| Regulatory regime | Description | In Scope (yes / no) |
| NERC CIP (US) | North American Electric Reliability Corporation critical infrastructure protection | NO |
| CRITICAL ELECTRIC (ENERGY) INFRASTRUCTURE INFORMATION (CEII) | US regulation covering specific engineering, vulnerability, or detailed design information about proposed or existing critical infrastructure | NO |
| NIS-D (UK) | EU NIS Directive/UK NIS Regulations 2018 set out cybersecurity obligations for network and information systems in the critical national infrastructure. | NO |
| FINANCIAL | Sarbanes-Oxley Act (SOX) is a federal law for auditing and financial regulations for public companies. | NO |
| HEALTH | Health Insurance Portability and Accountability Act (HIPAA) | NO |
| CREDIT CARD | Payment Card Industry Data Security Standard. PCI/DSS | NO |
| PRIVACY INCLUDING PERSONAL IDENTIFIABLE INFORMATION (PII) | EU GDPR (General Data Protection Regulation) and DPA 2018 (Data Protection Act) | YES (Customer and Worker PII master data) |
| Others |  |  |

## Baseline Security Requirements (BSR’s)

Link to [Base Security assessment](https://teams.microsoft.com/l/file/B8E39493-0D55-4569-9443-5DEF4E43DF93?tenantId=f98a6a53-25f3-4212-901c-c7787fcd3495&fileType=xlsm&objectUrl=https%3A%2F%2Fnationalgridplc.sharepoint.com%2Fsites%2FGRP-INT-Architecture%2FShared%20Documents%2FData%20and%20Information%20Architecture%2FData%20Foundations%20-%20Reltio%20%26%20Matillion%2FReltio%2FBaseline%20Security%20Requirements_v3.1%20-%20Reltio%2008-02-2021.xlsm&baseUrl=https%3A%2F%2Fnationalgridplc.sharepoint.com%2Fsites%2FGRP-INT-Architecture&serviceName=teams&threadId=19:0208706c699049ac805a806775adce1f@thread.skype&groupId=6e41877a-9b63-4354-b522-463045e111f7) of Master Data Management platform.

## Security Pattern

A pattern view of the target conceptual architecture, overlaid with BSR controls / references

Refer the architecture diagram, section

## Security Tower Engagement

This section applies to application teams and the Security Tower Engagement resulted in the Security BSRs.

## Additional Security Patterns

Reference [Base Security assessment](https://nationalgridplc.sharepoint.com/:x:/r/sites/GRP-INT-Architecture/Architectural%20Review%20Board/ITGC%20and%20SDA%20Sign%20Up/Archived%20ARB%20Submissions/2021/2021-01-19/Baseline%20Security%20Requirements_v2.6%20-%20INVP%205917E.xlsm?d=wc4ca033159614b058a2ebbe0f88798ff&csf=1&web=1&e=1Kl03d) v2.6 for INVP 5917E.

# IT Commercial

