Common Data Standards DRAFT

Feb 2022

Contents

[Purpose 2](#_Toc96694487)

[Context 3](#_Toc96694488)

[Documentation Requirements 4](#_Toc96694489)

[Conceptual Data Modelling Standards/Naming Conventions 7](#_Toc96694490)

[Logical Modelling Standards/Conventions 9](#_Toc96694491)

[Physical Modelling Standards/Conventions 13](#_Toc96694492)

[Naming Conventions for CamelCase platforms 15](#_Toc96694493)

[Packaged Solutions Standard & Requirement 19](#_Toc96694494)

[Existing System Standard & Requirement 20](#_Toc96694495)

[Addressing Security and Role-Based Access Control (RBAC) 21](#_Toc96694496)

[Data Asset Scanning Standards (EDC) 28](#_Toc96694497)

[Tagging of Data Assets: EDC and Cloud-based Platforms 30](#_Toc96694498)

[Review Process: Design and Pre-Deployment 31](#_Toc96694499)

[Required Artifacts 37](#_Toc96694500)

[Appendix 38](#_Toc96694501)

[Document Administration 39](#_Toc96694502)

Purpose

This document is designed to provide readers (specifically targeted to those designing and implementing data platforms for National Grid) with guidance to the Common Data Standards contained herein, as they are applied to the National Grid Customer Data Platform (CDP), and other efforts leveraging the shared technology stack (Snowflake, Matillion, Reltio MDM, and MuleSoft).

The reader is advised to read the Governance Program Overview before reading this artifact, as it provides context to the CDP Governance program, its actors, and requirements.

Context

Data governance (DG) is the process of managing the consistency, usability, integrity and security of the [data](https://searchdatamanagement.techtarget.com/definition/data) in enterprise systems, based on internal data standards and policies that also control data usage. The primary goal of an effective data governance program, is to ensure that the company’s investment into its data management environment is protected.

Data governance is everyone's responsibility, covering all areas from data integration, quality and MDM, to data modelling and warehousing.

All data teams and personnel (regardless of employment status: full time, independent contractor, and consulting firm employees, etc) are required to follow the Common Data Standards, in addition to any platform-specific standards that may apply.

All new systems/resources that are custom-built by/for National Grid must comply with all National Grid data, integration, and MDM standards.

Documentation Requirements

The general documentation structure for the data domain in the governance program is depicted below.



|  |  |
| --- | --- |
| **Artifact/Item** | **Comment** |
| Top Level – Governance Program Overview  (artifact comes in document and PPT formats) | Describes the structure of the overall (and easily extensible) governance program, its components, processes, tools, and deliverables.  Covers MDM (Reltio), ETL (Matillion), and Data.  Required reading for all personnel performing CDP delivery work at National Grid, regardless of employment status (contractor, consultant, FTE, etc.). |
| Common Data Standards | This artifact. Provides the data technologist with common standards that apply to all data platforms, program requirements (deliverables, processes), data tools, and best practices. |
| Platform-specific standards | Technology resources working in specific data platforms must read the platform-specific standards, as best practices, platform technical/organizational frameworks, and other site-specific information are detailed here.  Snowflake is the primary CDP data platform. Others can easily be added to the governance program-supported selection of platforms. |
| Common Data Design Template | This template contains a standardized selection of section headings with instructions included for each.  Includes Background, Requirements summary, decision made, references, diagrams, critical transaction designs, security considerations, and data platform organization, among other topics. |
| ER/Studio Repository. | Contains vital (required!) design artifacts, including:   1. Conceptual Data Models 2. Logical Data Models 3. Physical Data Models   This repository is to be regularly scanned by the Informatica Enterprise Data Catalog (EDC). This provides an automated mechanism to the EDC, and therefore Axon, to remain fresh.  The data design metadata (databases, schemas, objects, attributes, and comments) populates the EDC, which in turn matches design metadata with data resources and tools, to help automatically document the National Grid data infrastructure. |
| Delivery | Added for completeness: the DDL generated by ER/Studio is used to build/maintain custom-developed National Grid data platforms. |

The primary take-away is that a Data Design document must be included with every submission to the Governance Program (Data) Review.

* The documents are to be considered LIVING and must be maintained with every release.
* Large systems can use the template in a header-detail documentation model: describe the overall system (header), and then segregate the subsystems into dedicated artifacts.
* If your documents are approaching or exceeding 20 pages, consider breaking the documentation into smaller/easier-to-digest artifacts.
* Large systems might have to be reviewed in phases. Plan your work accordingly: lousy planning on the part of the database designer/architect does not necessarily constitute an emergency on the part of the data management team. This applies to the data access and application development teams as well.

Conceptual Data Modelling Standards/Naming Conventions

The Conceptual Data Model represents the high-level view for one or more data subject areas for a business area. It is a data model that is presented at a high level of abstraction, hiding the underlying details, and making it easier for people to comprehend.

A Conceptual Data Model should reflect the phenomena in the Users’ world being modelled as directly as possible, as close to the way the Users think, for example, many to many relationships are common in conceptual models.

This model is created and maintained by individual products, programs, and projects. It is a mandated artifact and must be created and kept up to date by them. It is critical that a CDM be created early in the product, program, or project scope as a work product to make sure it is aligned with the ECDM and to articulate the data needs. The CDM is the first step in understanding the data requirements, and therefore needs to be carried out during the Requirements Gathering and Analysis phase. Whether a project or program uses agile or waterfall methodology does not affect or reduce the need for an accurate CDM.

Conceptual Data Models should conform to the following guidelines, summarized in the table below.

|  |  |
| --- | --- |
| **Standard** | **Example/Comments** |
| Each entity must be labelled with a noun and must not be plural. | Company, Party, Payment |
| Each Entity must be named using mixed case with the first letter capitalized and the rest of the word lowercase only | Company, Party, Payment |
| Each relationship must be labelled with a Verb statement with the first letter of each word capitalized. | Is A Member Of, Drives The Work Performed By |
| Many to many relationships are acceptable. |  |
| Cardinality is not required, although early identification is preferred. | The sooner cardinality information is known, the more effective the analysis and physical design processes will become. Cardinality is required in the Logical Data Model. |
| Classifying entities may be represented either as hierarchies, or as supertype-subtype relations. |  |
| Large designs should be broken out into topical subject areas | This promotes ease of readability |
| Care should be taken to minimize crossing relationships between entities. | This promotes ease of readability and the ease of communication data relationships |
| Business rules, or special processing requirements, should be included as a text block on the Conceptual Diagram. | Not typically included, but if important rules/design considerations need to take place, documenting them as early in the process as they are known helps ensure a complete design. |

Please see the [National Grid Data Modeling Standards](https://nationalgridplc.sharepoint.com/sites/GRP-COMMS-Global-Architecture/Shared%20Documents/Forms/AllItems.aspx?id=%2Fsites%2FGRP-COMMS-Global-Architecture%2FShared+Documents%2FKnowledgeBase%2FInformation%2FNG+Data+Modeling+Standards+and+Guidelines+-+Latest+Version.pdf&parent=%2Fsites%2FGRP-COMMS-Global-Architecture%2FShared+Documents%2FKnowledgeBase%2FInformation&isSPOFile=1) document for more detailed information.

Logical Modelling Standards/Conventions

The purpose of the Logical Data Model is to represent the data/information requirements of the business. To identify the business data (Entities), how they are related (relationships) and the business data (attributes) that is required to support the business. A Logical Data Model (LDM) is constructed using an Entity Relationship Diagram (ERD). The LDM is at the detailed level necessary to identify, name, describe and associate the required data elements. It captures the Business Data Rules (not process rules). It uses the Subject Area CDM as a blueprint to define, at a lower level, elements of the model that cover details of the subject area or a business function shown in the Subject Area CDM. The LDM extends entities and relationships depicted in the CDM to include keys Data Modeling Standards and Guidelines June 2021 Page 13 and attributes. (What is the meaning here? Confusing) The LDM does not impose or consider any technical requirements or limitations. It is a business model.

These models are created and maintained by/for individual products, programs, and projects. It is a mandated artifact and must be created and kept up to date by them. The LDM is created during the early part of the Design phase. As more details become available during the business process definition, LDM’s go through a set of revisions until the full scope of project’s data requirements are covered within the model. Only after a business review and acceptance, should the logical model be converted into a first-pass physical model.

The ECDM and subsequent CDM are the guides for the LDM. The LDM has to reflect the data concepts and entities of the conceptual models.

Logical Data Models should conform to the following guidelines, summarized in the table below.

|  |  |
| --- | --- |
| **Standard** | **Example/Comments** |
| Each entity must be labelled with a noun and must not be plural. | Company |
| Each Entity must be named using mixed case with the first letter capitalized and the rest of the word lowercase only | Company |
| Each relationship must be labelled with a Verb statement with the first letter of each word capitalized. | Is Classified By |
| Many to many relationships must be resolved. | This ensures that the distinctness and purpose of the relationship is fully recognized and obvious to the reader. |
| Each entity must have all attributes fully specified, and they must be fully exposed in the Logical Data Model. | All attributes must be commented with describing the purpose of the attribute by the data platform designer.  This provides clarity to the reviewers, developers, and users of the platform. |
| Cardinality is required, although early identification is preferred. | Cardinality information is critical to physical design and capacity planning. |
| Attributes must be named with the first letter of each word capitalized. | Company\_Address |
| Classifying entities may be represented either as hierarchies, or as supertype-subtype relations. | Customer -> Individual Customer & Establishment Customer |
| Logical models will be resolved to 3rd normal form (3NF). |  |
| Each entity must use surrogate keys, and identify alternate keys (sometimes called natural or business keys) |  |
| Large designs should be broken out into topical subject areas | This promotes ease of readability |
| Care should be taken to minimize crossing relationships between entities. | This promotes ease of readability |
| Entities must reveal all of their attributes in the logical diagram (i.e. no hidden attributes). | 1. This promotes ease of readability 2. Reviewers HATE having to search for what should be readily visible. |
| Business rules, and/or special processing instructions for attributes, are to be included in the “Notes” section of the design tool. | This aids physical database designers and data engineers to fully understand the significance of a given feature/attribute. |
| Text blocks will be used to describe important topical/technical notes or special handling rules for subsystems, subject areas, and/or complex data structures. | This aids physical database designers and data engineers to fully understand the significance of a given subject area/sub-model/object collection, and the rules required to successfully implement them. |

Physical Modelling Standards/Conventions

General physical standards are presented in this section. The reader/data platform designer/engineer is strongly advised to read platform-specific standards for details/specifics on the platform in question.

**General Physical Design Rules/Standards**

|  |  |
| --- | --- |
| **Rule/Standard** | **Comments** |
| Prior to deployment to test, ALL data deliverables must undergo a pre-deployment review. | The CI/CD-Jenkins pipeline will not permit deployment to test/QA unless it has been approved by the reviewers.  The reader/modeler/architect/DB designer is advised to ensure their delivered DDL and related SW conforms to standards, and to the approved design.  If you find the approved design is not feasible/reasonable for any reason, consult the Data Modeling, DG and architecture team prior to proceeding/committing to the current path. |
| All physical data models will be authored and generated from the ER/Studio data design tool, prior to check-in to bitbucket. | All generated DDL must also include COMMENTS for objects tables and ~~attributes~~ columns. |
| On-premise DDL for traditional RDBMS/ORDBMS must include storage clauses for objects and indexes. | The Implementation Physical Model must include the primary key constraints along with the indexes. |
| DDL creation for all databases will be based on the Logical Data Model that has undergone a First Pass Physical and Implementation Physical model process. | The DDL will be generated by ER/Studio and generally by the data modeler after discussions with the DBA’s..  Once the DDL is implemented, a Synchronization process will occur in which the data modeler will reverse engineer the newly implemented database to ensure that the database follows the original design. |
| Critical business transactions (a business transaction may be comprised of one or more individual database transactions) must include a QUERY PLAN(s) demonstrating that the entire/collective business transaction has been correctly written and/or tuned. | Database transactions in the cloud area billed for every execution of each transaction.  Hence – the performance and efficiency of each transaction matters not only because we want our consumers to be satisfied with the performance of our platform – but also because it means saving money.  Wasted motion (IO, CPU utilization, DB transactions, etc.) in the cloud means *leaving money on the table for the cloud vendor*: subsequently, efficient design (applications, components, ETLs, report development, etc.) and coding must be on the minds of ALL personnel. |

Platform-specific requirements can be found in the documentation for the platform in question. If there are no platform-specific standards published for the platform in question, then you cannot use that platform as the basis for your design.

Naming Conventions for CamelCase platforms

Let’s discuss. This should be an exception not the rule.

For the purpose of this discussion, a “CamelCase” platform is one that by virtue of a derived industry standard uses a CamelCase naming convention. Examples of this include all JSON/document-based platforms, such as MongoDB and Cassandra.

All the same naming standards apply to namespaces/schemas, objects, and attributes – but in the case with CamelCase data platforms, the underscores are to be removed unless the vendor embeds them, and replaced with capital letters representing each defining aspect of the object/attribute identifier.

Acronyms that designate OWNERSHIP of an object will remain in all CAPS, as shown in the examples table below.

Examples:

|  |  |
| --- | --- |
| **Standard/Relational Convention** | **CamelCaseConvention** |
| Object Name: CDP\_Customer\_Address | CDPCustomerAddress |
| Attribute Name:  Customer\_Id | CustomerId |
| AttributeName:  City\_Name | CityName |
| All objects and attributes must be as fully defined as the targeted platform (and DB design tool) allow, including object and attribute comments/descriptions. |  |

For more information, please see the platform-specific standards for the data platform in question, or otherwise consult your project data architect.

# ER studio access process

There is a NEW process for requesting US domain accounts for the purpose of accessing ER Studio.  A US user must submit this SNOW request on behalf of the UK user.  [This request form](https://nationalgrid.service-now.com/ksp?id=ksp_cat_item_guide&sys_id=ff2430e40f5fb600ce8247ece1050e49) will trigger an automated process that creates a US domain (nmpc) account for the UK user linking the UK user's employee ID with the UK account.  The person submitting the request should specify in the comments that this nmpc account is secondary to the existing UK account.

The UK user does not need to log into his/her machine with this account; it is meant to only access ER Studio.

Once the account is created, the requester will receive an email with login information and instructions and should pass that information to the UK user along with the link to [Citrix XenApp - Logon](http://appportal/Citrix/Appportal/auth/login.aspx).  The UK user will follow the instructions for activating the account.  The UK user should also log into Citrix at least once per week to keep the nmpc account from going dormant.  If the NMPC account becomes dormant, the UK user will need to contact the help desk to reactivate it.  Note that dormancy processes are automated and are required as part of user account management.

Further information on the process of ER studio access can be found [here](https://nationalgridplc.sharepoint.com/:w:/r/sites/GRP-INT-Architecture/Shared%20Documents/Data%20and%20Information%20Architecture/DataModeling-ERStudio/ER%20Studio%20FAQ%202.docx?d=we0ce937926db40b684dbd47fa17deb3c&csf=1&web=1).

# Team Server Access in the UK

The first step currently is to get access to the US LDAP. This is a temporary process until Global LDAP is in place. Here’s how to do that:

The US domain accounts for the purpose of accessing ER Team Server have to be requested. A US user must submit this SNOW request on behalf of the UK user.

The persons to contact about submitting this request are (at the time of this writing)

1. Sucharitha (Suchi) Kalwala: Sucharitha.Kalwala@nationalgrid.com, or
2. Sean (Thomas) Mcauley: Thomas.Mcauley@nationalgrid.com

This request form will trigger an automated process that creates a US domain (NMPC) account for the UK user linking the UK user's employee ID with the UK account. The person submitting the request should specify in the comments that this NMPC account is secondary to the existing UK account.

The UK user does not need to log into his/her machine with this account; it is meant to only access ER Studio.

Once the account is created, the requester will receive an email with login information and instructions and should pass that information to the UK user along with the link to Citrix XenApp - Logon.

The UK user will follow the instructions for activating the account.

The UK user should also log into Citrix at least once per week to keep the nmpc account from going dormant.

If the NMPC account becomes dormant, the UK user will need to contact the help desk to reactivate it. Note that dormancy processes are automated and are required as part of user account management.

**How to log into ER studio Team Server:**

The link for Team Server is [here.](http://erstudio.nationalgrid.com/Login)

To login, enter the same User Name and Password as that for your Windows account; upon successful login, go to ER Tools and expand Projects.

Packaged Solutions Standard & Requirement

Packaged solutions, regardless on-premise or SaaS, often come with best practices, procedures, and conventions that govern the enhancement and extension of the platform.

Requirement:

1. All personnel working on these systems will follow the conventions and/or standards as recommended by the platform vendor.
2. The only exception to the above, are site-specific requirements, or practices, that have been determined/documented by the implementation team.

Existing System Standard & Requirement

If/when you are performing maintenance, or enhancing an existing system (regardless custom, SaaS, or packaged solution), you must adhere to the dominant naming/coding conventions present in the system.

You are NOT to create any new standard, methodology, or naming conventions, just because you don’t like how the work was originally done.

If the system is being newly developed, and the team is not entirely committed to the current path, and you believe you have a better idea: bring it to your project architect who will bring it to the Enterprise Information Architect responsible for the relevant standards.

Addressing Security and Role-Based Access Control (RBAC)

Following are details of roles based access controls (RBAC) in-place at National Grid and detail including links for current information.

All data platforms (and product-related schema's designed for general consumption) must conform to RBAC, with specific roles created for:

ETL

Services

Reporting

Administration

Interactive SQL

# [**Credential and Key Management**](https://nationalgridplc.sharepoint.com/sites/GRP-INT-US-SecurityBusinessPartneringIntegration/BPI/Credential%20and%20Key%20Management.aspx?CT=1641389359158&OR=OWA%2DNT&CID=9d0b94a7%2D96a1%2Dacfc%2Dc9a5%2D900228819b0a)

**1. Introduction**

​The purpose of this document is to provide guidance and direction with managing, obtaining and understanding the requirements supporting the Credential and Key management Baseline Security requirement for SaaS and PaaS solutions, including:​

* Single Sign-On with SaaS Providers
* If SSO is not an option, Adhere to NG Password Rules
* Review Administrative Accounts and Privileges
* ​API Key management​

**Key Pair authentication**

Prior to onboarding any 3rd party application you must perform due diligence on the vendor by requesting the ISO 27001 certification along with Statement of Applicability and a SOC II Type II report.​​​

If a Vendor Assurance questionnaire (see link below) was completed during the Commercial/Procurement process, Section A2.2- “Do you support integration with existing customer-based SSO? Do you support identity federation standards?” This will provide you with additional insight into the Vendor’s identity capabilities.

[National Grid Vendor Assurance Questionnaire](https://nationalgridplc.sharepoint.com/:x:/r/sites/GRP-INT-US-SecurityBusinessPartneringIntegration/Shared%20Documents/Security%20Requirements%20and%20Standards/BSR%20Guidance/Other%20Guidance/Vendor%20Assurance%20Questionnaire_National%20Grid%20October%202019%20v0.1.xlsx?d=w8406cf9949124f089866ff8215dcf271&csf=1&web=1&e=ltIuQf)​ ​

If your due diligence results are satisfactory, NG Active Directory IDP Federation and MFA SSO can be implemented. Lastly, the project team should coordinate a meeting with the correct resources within the Vendor’s organization.

**2. Key information**

**2.1**Single Sign-On with SaaS Providers​

The prerequisites for SSO include:​

* + List of NG Users who would access this SaaS Application
  + Onboard 3rd Party Application in SNOW​ ​

**​2.1.1**  Obtain AD Group name—list of NG Users-- from Project Team or create a request in SNOW to add users to a Group using this link:​​

[Add Users to AD Group](https://nationalgrid.service-now.com/ksp?id=ksp_cat_item&sys_id=f85e7%E2%80%8B00ddb3f67840bb3644a4b9619c0)

Please note, you must have the proper authority to add Users to an AD Group.

Additionally, the group name should start with DELL, discuss details with the AD team once SNOW request is submitted. This group will be mapped to the SAML authentication and MFA by the O365 admins. ​

**2.1.2** Prior to opening a SNOW request to onboard SaaS provider/Application, obtain the following metadata from the 3rd Party Vendor, which normally includes:

* IdP Issuer URI
* IdP Single Sign-On URL
* IdP Signature Certificate (Must be in .PEM or .DER format)
* Client user email domains for routing rule ​

**2.1.3** If SaaS solution and application is ONLY accessible from corporate network, SaaS provider should be asked to whitelist Z-Scaler IPs. This will prevent access from personal devices. Z-Scaler IP information can be obtained from this site:  <https://config.zscaler.com/zscloud.net/cenr>. The recommendation for US and UK whitelisting is:

#### Washington DC 104.129.194.0/23

#### New York III 165.225.38.0/23

#### Upcoming DC I 185.46.212.0/22

#### London III 165.225.80.0/22

#### Manchester 165.225.196.0/23

​

Testing is simply, enter URL from NG asset, and ensure SSO is working and access is granted to SaaS site. Try same URL from personal device, access should be denied. ​

NOTE: When exchanging Metadata with the vendor send it embedded in a Word document this will avoid our email servers to re-write the URL if directly sent by email, ask the Vendor to send this in protected way (i.e. Encrypted Email or protected Word document).

* ​​Open a SNOW request, select Azure AD Application Onboarding using the following link: [Azure AD Application OnBoarding Request](https://nationalgrid.service-now.com/ksp?id=ksp_cat_item&sys_id=321f4eb5db03c89039a29785ca961976)
* If they support SAML - select new enterprise app in the SNOW request.  If they​​ support Oauth/OIDC - select new application registration in SNOW request
* Upload the vendor's metadata file​
* Have the group name created in above process available.
* Request for the metadata to be sent to you once the ticket is completed.

Please refer to Cyber Security Services (Product) and Cyber Security Services (Engineering)Team organization charts for IAM contacts.

[National Grid Security Team Org Chart](https://nationalgridplc.sharepoint.com/sites/NationalGridSecurity/Shared%20Documents/Forms/AllItems.aspx?id=/sites/NationalGridSecurity/Shared%20Documents/NG_Security_Organisation_Chart.pdf&parent=/sites/NationalGridSecurity/Shared%20Documents&p=true&originalPath=aHR0cHM6Ly9uYXRpb25hbGdyaWRwbGMuc2hhcmVwb2ludC5jb20vOmI6L3MvTmF0aW9uYWxHcmlkU2VjdXJpdHkvRVdCRWs3Tk44U3BIakZXRlRJVmxQU1FCNlhPXzZTWkRWV2xCTHJqbFh1S0dNUT9ydGltZT1DODdkcFR1aDJFZw)

​​

**3.0 Adhere to the NG Password rules if not integrated with a NG SSO service.**

If the SaaS provider does not support SSO, you can reference the Vendor Assurance Questionnaire (if one was obtained) -- Link provided above -- item # A2.5 "Please describe how your password parameters configuration meets National Grid security requirements."

Additionally, you can reference National Grids Global IT Control Framework, which can be located here:

[Security Standards, Policies and Procedures](https://nationalgridplc.sharepoint.com/sites/NationalGridSecurity/SitePages/Policies-%26-Standards---Business-Management-System.aspx)

National Grid's password restrictions can be found under Identification & Authentication, Control # IDA-4. Evidence of the SaaS provider's password policy should be obtained (If SSO integration is not supported)​

**4.0 Review Administrative Accounts and Privileges quarterly**

If a SOC 2 Type II has been provided, assurance regarding the SaaS provider's Periodic Access Review, System Account Management, or Monitoring Activity controls can be obtained for reference.

If a SOC 2 has not been provided, or the Vendor does not have a SOC 2 Type II, request copies of their Security policies and standards

**5.0 API Keys**

API keys provide a simple mechanism for authenticating applications. API Key Management verifies calls from applications or sites requesting access to an API, validating and permitting only those with valid keys. There are 2 API keys (tokens), one for ID and one for access.

First, determine if APIs will be used on the project between National Grid and the SaaS provider for data transfer. If APIs are to be used, follow the process outlined below to obtain ID and access tokens.

Please note the same process for Onboarding an application into Azure AD will be used:

[Azure AD Application OnBoarding](https://nationalgrid.service-now.com/ksp?id=ksp_cat_item&sys_id=321f4eb5db03c89039a29785ca961976)​

When asked in SNOW request to “Please select Token Type” select Both access Token and ID token for your APIs​.​

**6.0 Contacts**

Currently, the following individuals are supporting IAM at the time this document was created.

* IAM Delivery: Dedat-Humphrey, Nasim:  [Nasim.Dedat-Humphrey@nationalgrid.com](mailto:Nasim.Dedat-Humphrey@nationalgrid.com)
* IAM Engineering: Tammy Cooper: Tammy.Cooper@nationalgrid.com
* IAM Product Owner: Plessas, John:  John.Plessas@nationalgrid.com​

For additional assistance with this Baseline Security Requirement, please contact your Secuity Architect for your project​​

Please refer to the following link

<https://nationalgridplc.sharepoint.com/sites/GRP-INT-US-SecurityBusinessPartneringIntegration/BPI/Credential%20and%20Key%20Management.aspx?CT=1641389359158&OR=OWA-NT&CID=9d0b94a7-96a1-acfc-c9a5-900228819b0a>

**7.0 Additional Notes**

Please refer to AAA-5 (Two Step Authentication) for accessing external applications from outside of National Grid corporate network

Source from “[Credential and Key Management](https://nationalgridplc.sharepoint.com/sites/GRP-INT-US-SecurityBusinessPartneringIntegration/BPI/Credential%20and%20Key%20Management.aspx?CT=1641389359158&OR=OWA%2DNT&CID=9d0b94a7%2D96a1%2Dacfc%2Dc9a5%2D900228819b0a)”

<https://nationalgridplc.sharepoint.com/sites/GRP-INT-US-SecurityBusinessPartneringIntegration/BPI/Credential%20and%20Key%20Management.aspx?CT=1641389359158&OR=OWA-NT&CID=9d0b94a7-96a1-acfc-c9a5-900228819b0a>

# National Grid Baseline Security Requirements (BSRs) Index with ​​​​​​​​links to individual BSR Guidance can be located here:-

<https://nationalgridplc.sharepoint.com/sites/GRP-INT-US-SecurityBusinessPartneringIntegration/BPI/BSR%20Guidance.aspx?CT=1641893799436&OR=OWA-NT&CID=c5508a49-14a8-7d0f-5027-96cb1e19028e>

# **Informatica User & Security Management**

Informatica tooling (EDC/Axon/CDQ) supports three options to configure roles based user access: natively (defined within Informatica platform, with LDAP integration or SSO/SAML access.

**LDAP INTEGRATION**

Can be configured as an lnformatica domain to enable users imported from an

LDAP directory service to log in to lnformatica nodes, services, and

application clients

Axon can connect to the LDAP directory in your organization, retrieve users,

assign them user profiles, and display them in the Axon interface. Axon

automatically creates the users after it connects to the LDAP server

LDAP providers certified: MS AD, OpenLDAP etc.

Refer to PAM for compatibility

**SSO (SAM L)**

You can configure Security Assertion Markup Language (SAML) authentication in an lnformatica domain and in Axon installation.

For lnformatica domain, we support Microsoft Active Directory Federation

Services (AD F8] identity provider

For Axon, we certify Okta, OneLogin and Azure Active Directory. However,

Axon does support with all vendors supporting SAML v2.l],

Refer to PAM for compatibility

Source from “[Informatica](https://nationalgridplc.sharepoint.com/sites/GRP-INT-US-SecurityBusinessPartneringIntegration/BPI/Credential%20and%20Key%20Management.aspx?CT=1641389359158&OR=OWA%2DNT&CID=9d0b94a7%2D96a1%2Dacfc%2Dc9a5%2D900228819b0a) Website”

<https://www.informatica.com/content/dam/informatica-cxp/techtuesdays-slides-pdf/Deployment%20Best%20Practices%20of%20Data%20Governance%20Products-Axon,%20EDC,%20IDQ,%20DPM.pdf>

National Grid have implemented Axon using the LDAP standard and EDC using the SSO standard.

Data Asset Scanning Standards (EDC)

The Informatica Enterprise Data Catalog creates inventories of National Grid Data management assets (“resources”), and helps users search, understand, classify, and research the large volumes of metadata (and therefore systems) collected for/by the enterprise.

Physical and operational metadata can be extracted for many objects, organize the metadata based on business concepts, and view the data lineage and relationship information for each object. Enterprise Data Catalog (EDC) maintains a catalog. The data catalog serves as a centralized repository that stores all the metadata extracted from different external sources.

All technical metadata sources (data platforms, integration tools, reporting tools/environments, etc.) are required to be scanned into the EDC.

General requirements for information to include:

* English name of system
* Description of system
* Technical designation (i.e. name of data asset, such as Oracle SID)
* URL (if applicable)
* Location (on-premise, cloud)
* Business owner/steward of system (SME)
* Technical owner of system
* Platform type: DB (Snowflake, Oracle, SQL Server), ETL tool (Matillion, etc), Reporting Environment (Tableau, Power BI, etc.), Design Repository (ER/Studio, Power Designer, etc.), Azure BLOB, SaaS, etc.
  + See the complete list of resources that can be scanned in the Appendix section of this artifact.
* Username
* Password
* Port Number
* Environment (Dev, Prod, etc.)
* Schema(s) to be scanned
* Recommended scanning frequency (daily, weekly, monthly, per release, etc.)
* Department
* Restricted Access (Yes or No).
* Does the system contain PII, SPII, or CNI?

The process begins with a request completed for the new source system, and filling all the details of their system including account credentials. As needed, DG team will contact the originators of the request, and/or arrange meetings with source system data owners (and/or Steward). Once all security approvals are obtained, DG team to start working on the configuration of the scanner, if it is not already built and be able to scan the system to produce profiling results.

Tagging of Data Assets: EDC and Cloud-based Platforms

All cloud-based data assets (including all cloud-based data integration tools) are required to be tagged, and the tags used in the cloud must match the tags in the Enterprise Data Catalog (which also requires tagging of on-premise data platforms).

|  |  |
| --- | --- |
| **Tag** | **Description/Comment** |
| DBName | English Name of database |
| DBIdentifier | This would be the technical name of the database itself.  Example: Oracle SID |
| DBPlatform | Examples: Snowflake, Oracle, SQLServer, PostgreSQL, MongoDB, AzureBLOB, etc. |
| DBSchema | If the DB is a shared data asset, then add the schema |
| DBType | Warehouse, LakeHouse, OLTP, Repository, Graph, Name/Value, InMemory, etc. |
| Department | Department that owns the data asset |

.

Review Process: Design and Pre-Deployment

The governance program requires that all deliverables (MDM, Data, and ETL) undergo two reviews prior to proceeding to testing, and these are organized as follows:

1. Design review
2. Pre-deployment review

The design review provides the opportunity for the project architect, or the agile team data lead, to present their designs to the reviewers. These will be reviewed based on the quality of the design, adherence to standards, and completeness of the solution.

Note that logical and physical design metadata will be scanned on a regular basis from the ER/Studio repository into the Enterprise Data Catalog. Hence – data stewards, owners, your fellow architects, agile team leads, product managers, and other interested parties will have the ability to monitor all modelling work, and monitor adherence to standards.

The review process is depicted in the diagram below.



The design review must be scheduled at least 3 days in advance, and all deliverables must be included in the review request submission (this will be done in Service Now). For data, the required artifacts are described in the table below.

**Design Review Requirements**

|  |  |
| --- | --- |
| **Artifact** | **Description** |
| Logical Data Model | The logical data model must conform to the Logical Data Model standards summary elsewhere in this document, and/or otherwise the *National Grid Data Modelling Standards and Guidelines.*  Note that the architect/team data lead should be prepared to discuss the physical design approach and/or intent. |
| Data Design Document | This document requires use of the Data Design Document Template, which contain sufficient content that will be *targeted to a senior technical resource/engineer*, which is defined as “trained in the relevant technologies, with approximately 5-7 years of experience.”  The idea is that after reading the document and reviewing the data model(s), the individual (assuming they are already versed in NG data governance standards, etc.) should be able to be given a problem to solve, and know how/where to get started.  The goal is to protect the investment National Grid is making in their data management assets. |
| Supplemental Design Diagrams | These will be provided by the architect or agile team lead. This should include:   1. System Context Diagram – outlining where this system fits into the National Grid System landscape. 2. Data Flow Diagram(s), how data flows in and out of the system in question, including all data sources/destinations. 3. Process flows for critical transactions. 4. Database layout, depicting the schemas (or namespaces) in a given data platform, or a topology for Azure BLOB, etc., that describes the major components of the platform in question.   The above will be integrated into the Data Design Document. |

The pre-deployment review is a final check prior to releasing software deliverables for testing. No software will be promoted to testing without undergoing the pre-deployment review. The CI/CD-Jenkins pipeline depends on the Pre-Deployment Review “Passed Flag” to be set by the reviewer.

The pre-deployment review exists to ensure that:

1. The design that was approved is the same design that was implemented by the development team.
2. Naming standards are being adhered to
3. Coding standards are being adhered to, and that critical transactions have been reviewed for performance and clarity.

**Pre-Deployment Review Requirements**

|  |  |
| --- | --- |
| **Artifact** | **Description** |
| Physical Data Model | Both the First Pass physical data model and the Implementation Data Model ~~The physical data~~ model must conform to the Physical Data Model standards as required in the platform-specific standards for the data platform in question (i.e. Oracle, Snowflake, MongoDB, Azure BLOB, etc.)*.* |
| Data Design Document | This document requires use of the Data Design Document Template, which contain sufficient content that will be *targeted to a senior resource*, which is defined as “trained in the relevant technologies, with 5-7 years of experience.”  Note: The details behind physical schema design, indexing strategy (if relevant), and critical transactions will be detailed in this document, in addition to security considerations.  Again: the idea is that after reading the documentation and reviewing the data model(s), the individual (assuming they are already versed in NG data governance standards, etc.) should be able to be given a problem to solve, and know how/where to get started. |

Required Artifacts

Appendix

This section contains references to related documents and websites, etc.

[Informatica Data Governance](https://www.informatica.com/products/data-governance.html)

[Idera ER/Studio](https://www.idera.com/products/er-studio/enterprise-data-modeling)

[Reltio MDM](https://www.reltio.com/)

[Matillion ETL/ELT](https://www.matillion.com/)

[BitBucket](https://bitbucket.org/product)

[National Grid Data Modeling Standards](https://nationalgridplc.sharepoint.com/sites/GRP-COMMS-Global-Architecture/Shared%20Documents/Forms/AllItems.aspx?id=%2Fsites%2FGRP-COMMS-Global-Architecture%2FShared+Documents%2FKnowledgeBase%2FInformation%2FNG+Data+Modeling+Standards+and+Guidelines+-+Latest+Version.pdf&parent=%2Fsites%2FGRP-COMMS-Global-Architecture%2FShared+Documents%2FKnowledgeBase%2FInformation&isSPOFile=1)

[Snowflake Security Model](https://confluence.us.ngridtools.com/display/CDTTEAM/Snowflake+-+Security+Model)

[Snowflake Operations Model](https://confluence.us.ngridtools.com/display/CDTTEAM/Snowflake+-+Operations+Model)

[CDP Snowflake Architecture and MVPs](https://confluence.us.ngridtools.com/display/CDTTEAM/Snowflake+Architecture+-+MVP1+Target+State+Architecture)

[Snowflake](https://www.snowflake.com/)

[Enterprise Data Management Framework](https://nationalgridplc.sharepoint.com/sites/GRP-INT-UK-DataManagement)

Document Administration

|  |  |
| --- | --- |
| Owner | Ken Jones |
| Author/s first edition | Christopher Twombly |
| Date first created | 25 February 2022 |
| Date of current edition | 25 February 2022 |
| Date last amended | 25 February 2022 |

Reviewers

|  |  |  |
| --- | --- | --- |
| Review Name | Role | Date & version |
|  |  |  |
| Ken Jones | Head of Data Governance | V0.1 DRAFT |
| TBD | Data Governance Senior Manager | V0.0 |
| TBD | Group Head of Data & Analytics | V0.0 |
|  |  |  |

Change History

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Author | Description of Changes |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

National Grid plc

National Grid House,

Warwick Technology Park,

Gallows Hill, Warwick.

CV34 6DA United Kingdom

Registered in England and Wales

No. 4031152

nationalgrid.com

National Grid plc

National Grid House,

Warwick Technology Park,

Gallows Hill, Warwick.

CV34 6DA United Kingdom

Registered in England and Wales

No. 4031152

nationalgrid.com