General Data Standards

xxxxx – The Principles and Standards

Jan 2022

Contents

[Purpose 2](#_Toc92461457)

[Context 3](#_Toc92461458)

[Conceptual Data Modelling standards/naming conventions 4](#_Toc92461459)

[Logical modelling standards/naming conventions 5](#_Toc92461460)

[Naming conventions for CamelCase platforms 6](#_Toc92461461)

[Platform specific design/modelling, naming conventions, and practices for Snowflake 7](#_Toc92461462)

[Packaged Solutions Standard & Requirement 14](#_Toc92461463)

[Data Asset scanning standards (EDC) 19](#_Toc92461464)

[Business glossary (Axon) 20](#_Toc92461465)

[The General Data Principles 23](#_Toc92461466)

[The General Data Standard Requirements 25](#_Toc92461467)

[National Grid Document Taxonomy 33](#_Toc92461468)

[Document Administration 35](#_Toc92461469)

Purpose

This document is designed to provide readers with guidance to General Data standards within National Grid, Principles and Requirements.

It is suitable for all employees who deal with data either directly or indirectly.

This document contains the following sections;

* The purpose of the guidelines and the target audience
* Introduction to General Data Standard & Principles
* A list of the General Data Standard Requirements
* Document Administration

Context

Data governance (DG) is the process of managing the availability, 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. Effective data governance ensures that data is consistent and trustworthy and doesn't get misused.

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

All data teams 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 National Grid must comply with all National Grid data, integration, and MDM standards.

Outlined in this document are the development standards relating to the following sections:-

* *Conceptual Data Modelling standards/naming conventions*
* *Logical modelling standards/naming conventions*
* *Naming conventions for CamelCase platforms*
* *Platform specific design/modelling, naming conventions, and practices for Snowflake*
* *Packaged Solutions Standard & Requirement*
* *Data Asset scanning standards (EDC)*
* *Business glossary (Axon)*

This lists the scope and sections of the general data standards document.

Conceptual Data Modelling standards/naming conventions

TheConceptual 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.

\*\*\*\* Insert Conceptual data modelling naming conventions here

For reference further detail on data modelling can be found here:-

\*\*\*\* Insert link to DMG 10 data modelling standards document here

Logical modelling standards/naming conventions

The Logical Data Model includes the more detailed view of data including keys, attributes and definitions of business objects, including the relationships between them.

It is an entity-relationship data model including data attributes that represents the inherent properties of the data, including names, definitions, structure and integrity rules, independent of software, hardware, volumetric, frequency of use or performance considerations.

\*\*\*\* Insert Conceptual data modelling naming conventions here

For reference further detail on data modelling can be found here:-

\*\*\*\* Insert link to DMG 10 data modelling standards document here

Naming conventions for CamelCase platforms

Naming conventions for CamelCase platforms (JSON, Streaming solutions, etc.).

\*\*\*\* Insert Camelcase info (from Maria) naming conventions here

Platform specific design/modelling, naming conventions, and practices for Snowflake

All relational data structures, JSON documents used for data storage, etc., must be catalogued/stored in ER/Studio, in accordance with the ER/Studio

Metadata conventions/standards.

-> See Maria

All DDL must be generated from the Idera ER/Studio repository;  All metadata (i.e. descriptive comments/definitions) must be deployed into data platforms

that support it.

->  See Maria for details, and instructions for generating DDL that can be reviewed by the architect/engineer

->  See Maria for details regarding how to get an ER/Studio licence (along with requirements/roles of those who need to use it), and for links to the

ER/Studio Portal

There are standards that we need to follow, when naming a component in snowflake or Matillion. Below are the naming standards for both Matiliion and Snowflake.

# **1.Snowflake:**

| **Database Components** | **Naming Convention** | **Example** |
| --- | --- | --- |
| Landing zone Schema | RAW\_SourceSystemName\_ProjectName | RAW\_CSS\_CDP, RAW\_CRIS\_CDP |
| Landing zone Table (Staging) | STG\_SourceSystemTable\_ | STG\_ |
| Landing Zone Table (Transformed table) |  | Same as source table names |
| Target Schema | SourceSystemName\_ProjectName | CDD\_CDP, CRIS\_CDP |
| Target table (Final) |  | Same as source table names |
| Views | VW\_Purpose\_Or\_Functionality | VW\_XXX |
| Procedures | USP\_Purpose\_Or\_Functionality | USP\_XXX |
| Roles | Env\_ProjectName\_ | DEV\_CDP\_ADMIN\_ROLE |
| Warehouse | Env\_ProjectName\_ | DEV\_CDP\_SUPPORT\_WH |
| Service Account | SVCProjectnameExternalapplicationname\_ENV | SVCCDPCDQ\_DEV |
| Database name | ENV\_ProjectName\_ | DEV\_CDP\_DEVELOP\_DB |
| Storage Integration | ProjectNameENV\_SI | CDPTEST\_SI |

# **2.Matillion:**

| **Matillion Jobs** | **Naming Convention** | **Example** |
| --- | --- | --- |
| **Main Job** | JOB\_XXX |  |
| CSS\_TO\_SF\_LOAD\_WF |  | JOB\_CSS\_TO\_SF\_LOAD\_WF |
| CSS\_TO\_SF\_MAIN\_WF |  | JOB\_CSS\_TO\_SF\_MAIN\_WF |
| CRIS\_TO\_SF\_LOAD |  | JOB\_CRIS\_TO\_SF\_LOAD\_WF |
| CRIS\_TO\_SF\_MAIN |  | JOB\_CRIS\_TO\_SF\_LOAD\_WF |
| JOB\_LOAD\_CSS\_CODE\_MAIN |  | JOB\_LOAD\_CSS\_CODE\_MAIN\_WF |
| JOB\_LOAD\_CSS\_CODE\_EXTRACT |  | JOB\_LOAD\_CSS\_CODE\_EXTRACT\_WF |

| **Component** | **Prefix** | **Example** |
| --- | --- | --- |
| Create Table | CT | CT\_XXX |
| Alter Table | AT | AT\_XXX |
| Delete Table | DT | DT\_XXX |
| SQL Script | SS | SS\_XXX |
| Database Query | DBQ | DBQ\_XXX |
| Excel Query | EXQ | EXQ\_XXX |
| Google Sheets Query | GSQ | GSQ\_XXX |
| Start | START | START\_XXX |
| End Failure | END\_FAILURE | END\_FAILURE |
| End Success | END\_SUCCESS | END\_SUCCESS |
| File Iterator | FLI | FLI\_XXX |
| Fixed Iterator | FXI | FXI\_XXX |
| Grid Iterator | GRI | GRI\_XXX |
| Loop Iterator | LPI | LPI\_XXX |
| Table Iterator | TBLI | TBLI\_XXX |
| Bash Script | BS | BS\_XXX |
| Python Script | PS | PS\_XXX |
| Alter Warehouse | AWH | AWH\_XXX |
| Create File Format | CFF | CFF\_XXX |
| Data Transfer | DXFER | DXFER\_XXX |
| Table Input | TBLIP | TBLIP\_XXX |
| Table output | TBLOP | TBLOP\_XXX |
| Calculator | CAL\_ | CAL\_XXX |
| Truncate Table | TRNCT\_ | TRNCT\_XXX |
| **Component** | **Prefix** | **Example** |
| **Component** | **Prefix** | **Example** |
| Begin | BEGIN | BEGIN |
| Commit | COMMIT | COMMIT |
| Rollback | ROLLBACK | ROLLBACK |
| Run Orchestration | ROR | ROR\_XXX |
| Transformation Orchestration | TOR | TOR\_XXX |
| Append To Grid | ATG | ATG\_XXX |
| Query Result To Grid | QRTG | QRTG\_XXX |
| Query Result To Scalar | QRTS | QRTS\_XXX |
| Remove From Grid | RFG | RFG\_XXX |
| Table Metadata To Grid | TMDTG | TMDTG\_XXX |
| JDBC Incremental Load | JDBCILD | JDBCILD\_XXX |
| Azure Blob Load Generator | AZBLG | AZBLG\_XXX |
| Salesforce Incremental Load | SSFIL | SSFIL\_XXX |
|  |  |  |

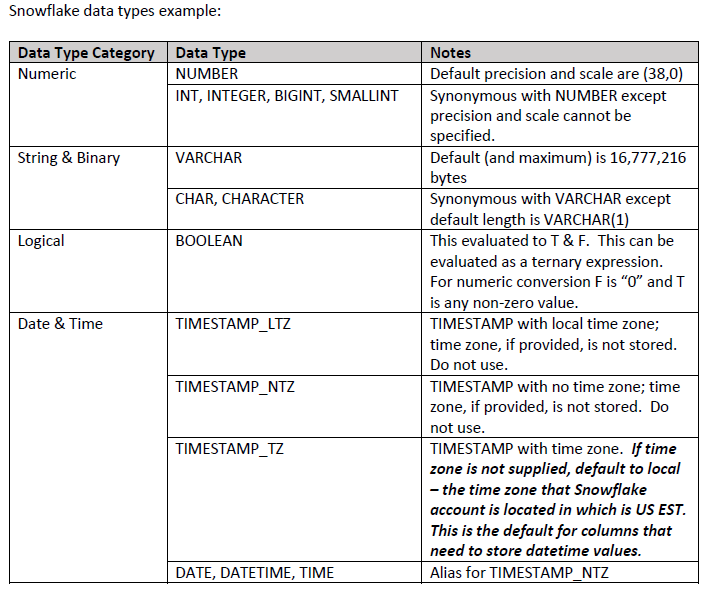
| **Type** | **Prefix** |
| --- | --- |
| Environment Variables | ENV |
| Job Variables | JOB |
| Grid Variable | GRID |
|  |  |
|  |  |

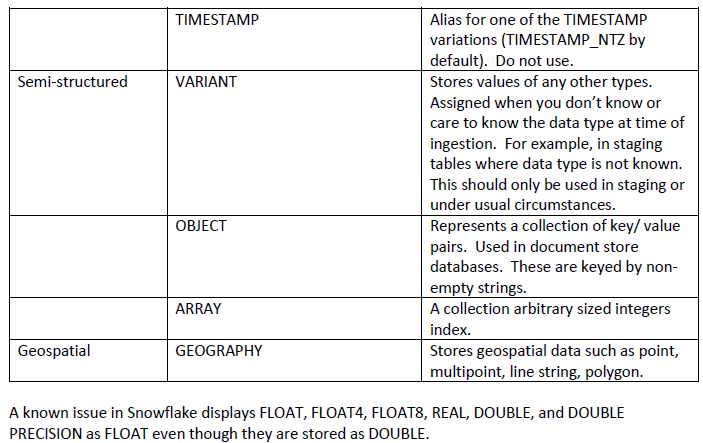
Source:-

<https://confluence.us.ngridtools.com/pages/viewpage.action?spaceKey=CDTTEAM&title=Naming+Conventions>

# [**DataType Convention**](https://confluence.us.ngridtools.com/display/CDTTEAM/DataType+Convention)

| **Datatype in Source** | **Datatype in Snowflake** |
| --- | --- |
| Integer(x,0) | Number(38,0) |
| Date(x,0) | TIMESTAMP\_TZ |
| Decimal(x,0) | Number(38,0) |
| char(x,0) | VARCHAR(x,0) |
| Timestamp(x,x) | TIMESTAMP\_TZ |
| BOOLEAN | BOOLEAN (TRUE,FALSE, 1 or 0) |
|  |  |





Source from "NG Data Modelling Standards and Guidelines"

link : <https://nationalgridplc.sharepoint.com/:b:/r/sites/GRP-COMMS-Global-Architecture/Shared%20Documents/KnowledgeBase/Information/NG%20Data%20Modeling%20Standards%20and%20Guidelines%20-%20Latest%20Version.pdf?csf=1&%3Bweb=1&isSPOFile=1>

Packaged Solutions Standard & Requirement

All contractors and consulting firms are required to comply with National Grid Data, Integration, and MDM standards, policies, and conventions.

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​

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**](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>

\*\*\*\* Insert active directory conventions here from Reza

# **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 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

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

Data Asset scanning standards (EDC)

\*\* Needs EDC standards inserting

All (each) EDC and Cloud/Azure Data assets must be tagged (Tagging Requirement!) as follows (with identical values):

Platform Owner: <NG Business Unit>

Platform Type,  one of

- SQLSERVER

- SNOWFLAKE

- AZUREBLOB

- POSTGRESQL

-> Document Store (ex:  MongoDB)?

-> In-Memory DB (ex: Redis)?

-> Name/value DB?

-> Graph?

-> Add Other NG-approved data platforms (meet with Fred Wright, Maria, GridStack Team, etc.)?

Data Resource Name: <Business DB Title/Name>

SchemaNamespace: <schema list>

- Others?  See Rama, Fred Wright, GridStack contact.

Business glossary (Axon)

Provide link to Informatica Axon.

All

Processes:

- Data Project Reviews

-> Design

-> Pre-deployment

- New Data Platform Request

- Data Design Dispute Arbitration

Others?

Appendix:

- Provide links to all policies, supporting materials

Ignore:----

General Data is a critical component in our digital transformation. It is a way of understanding our data and sharing that knowledge.

It requires a commonality of purpose and approach to maximize its impact and ensure positive outcomes.

Today, we’re focusing on the journey to a common General Data approach. It’s a combined effort of IT and business partners, both technical and non-technical.

The Intended Audience consist of data architects and modelers, data managers and leads, data analysts, business and systems, analysts, domain architects, data stewards, design and management and anyone interested in data.

This is primarily focused on discussing the new General Data standards. This will be a higher level than basic General Data training.

#

A definition of General Data is that it is “the process by which data (facts and statistics) is acquired, validated, stored, protected and processed; and by which its accessibility, reliability and timeliness is ensured to satisfy the needs of the data users”.  
  
Data are these facts, figures, numbers text etc. without context. e.g.10,000 feet. When data is added to a context it becomes information which can drive actions and, eventually, wisdom or insight.

Information is data that is:

* Accurate and timely,
* Specific and organised for a purpose,
* Presented within a context that gives it meaning and relevance.

The quality of our information is important to us because:

* It is how we represent all our assets (physical, financial and people)
* It underpins all the decisions we make
* It underpins our approach to risk Modelling
* It is how we demonstrate compliance
* It underpins the basis of how we are regulated, incentivised and how we make our returns on our assets

Therefore, General Data is the business function of effectively managing that data such that it becomes useful information. Through the effective implementation of a General Data Framework we can sustainably drive data integrity, accessibility, and value from our data.

National Grid has identified Data as one of our most important assets and taken the approach to embed General Data into the business as a core capability, enabling our data to be exploited as a key strategic asset, underpinning all of our key decisions and informing our future strategy.

Through the continuous and correct implementation of Data Modelling, and with support of our leadership, we can reach the desired level of maturity and trust in our data, which is critical to drive our vision and strategy for successful business outcomes

## The General Data Standard

National Grid has identified General Data as one of a number of core Standards created to support National Grid's Business Modelling System (BMS). General Data is included because:

“Data is one of our most important assets. The effective Modelling of data is essential to the delivery of safe, seamless and efficient services to our stakeholders. Robust data provides the basis for informed decision making and enables us to measure and improve performance, underpinning the delivery of our strategic objectives.”[[1]](#footnote-2)

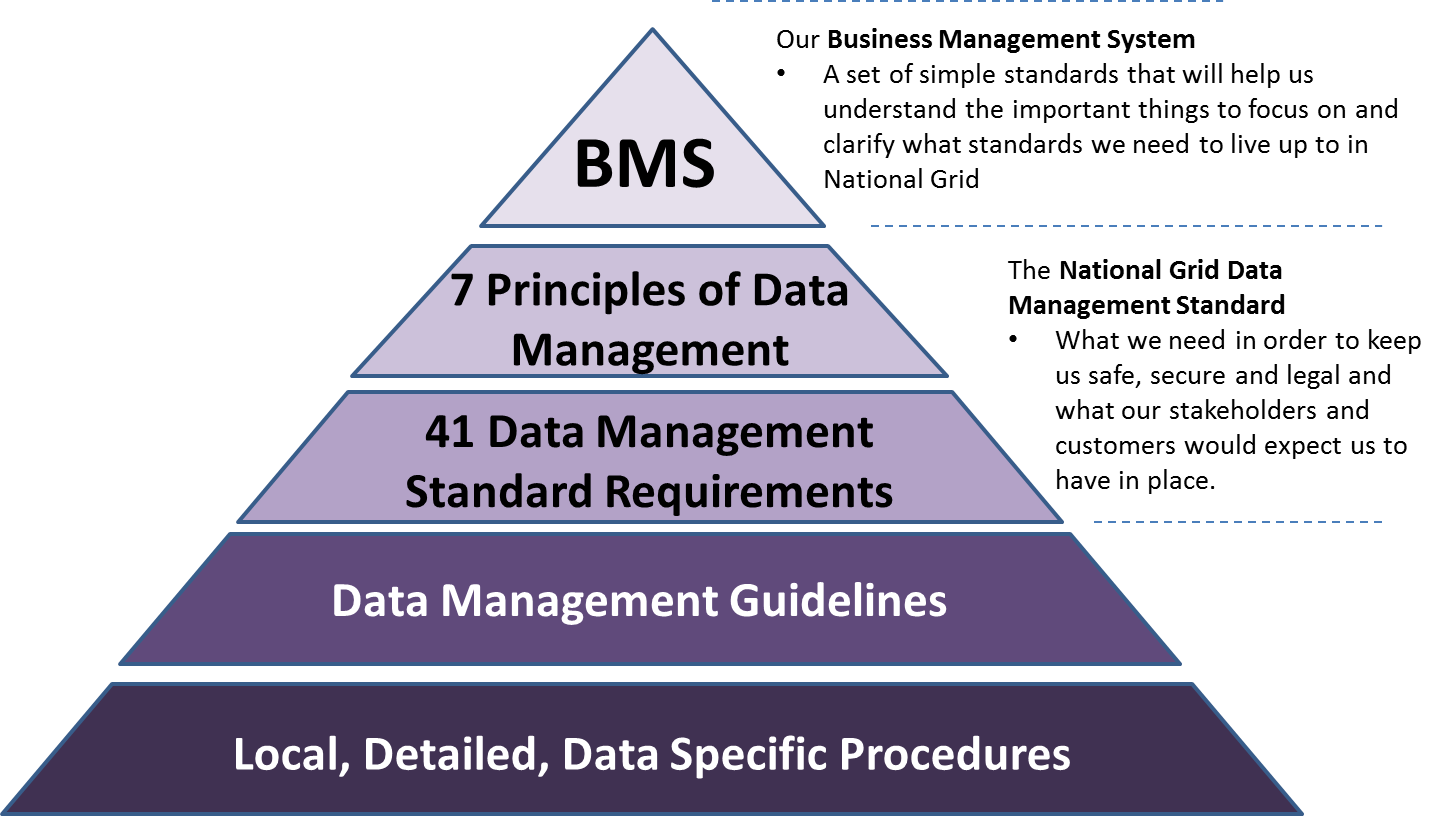
The diagram below illustrates what the implementation of this Standard will mean for us and what it will enable us to do:



The General Data Principles

In support of the General Data ambition, we have developed 7 General Data Principles supported by 41 Standard Requirements (previously known as Minimum Standards or Standards and updated in line with other, subsequently published, Standards) that set out what needs to be in place to meet The Standard and that National Grid has (globally) committed to achieving.

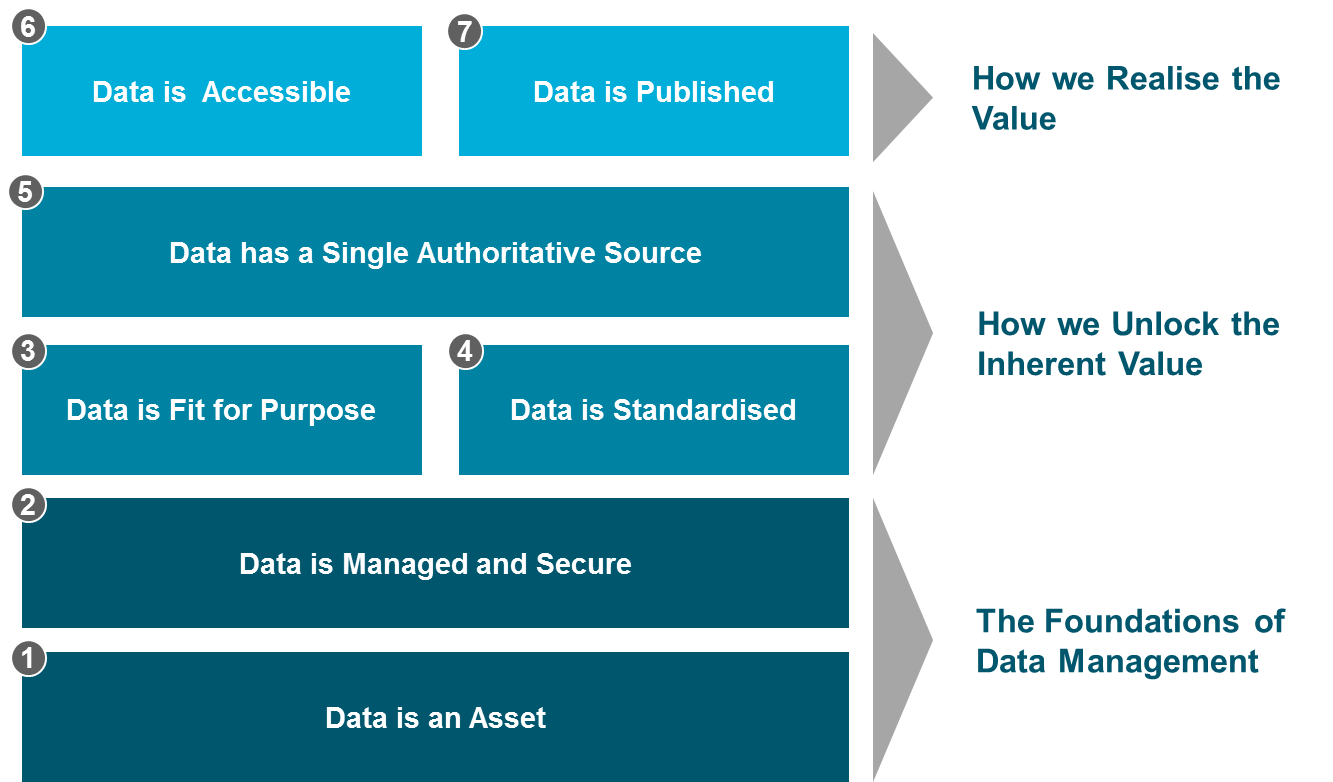
### Diagram to show the General Data components in Context:



The General Data Principles are intended to express a set of enduring themes to which all of our businesses and functions can commit to:

* These Principles apply to all data that is created, collected, held, used, shared, transformed, published, or processed by National Grid. They apply to both structured data (i.e. organised and searchable) and unstructured data (i.e. not organised and not readily searchable) at all stages of its lifecycle, including any historical data that is utilised by the business
* Each Principle consists of several parts covering the name and brief description, the rationale, the implications and the requirements
* Seven Principles have been defined and these builds naturally into a hierarchy (shown in the next diagram), with each principle building on what has gone before
* A set of supporting guidelines (including this one) have been produced to support the delivery of / compliance with The Standard.

### Diagram to show the 7 Principles of Data Modelling;



* The first two Principles ([1] & [2]) provide a foundation on which all of the others depend. It is important that our data is valued as an asset and managed, protected and exploited throughout its lifecycle. Data must also be governed, assured and secure from unauthorised access with ownership and other key organisational roles and responsibilities clearly defined and in place.
* The next two Principles ([3] & [4]) help unlock the value inherent in our data. Our data does not need to be perfect, but it must be fit for purpose, in terms of conforming to clearly defined quality characteristics. Our data also becomes inherently more valuable when it is made available in standardised formats and linkable to other data sources.
* With these pre-requisites in place, the Principle of re-use ([5]) with a single authoritative (master) source of data, can be achieved.
* The top layer of Principles ([6] & [7]) build on all the layers below, providing transparency by opening appropriate access to accurate and complete data to both business users and external stakeholders.

The General Data Standard Requirements

The General Data Standard, Principles, Requirements and Guidelines, cannot in themselves change the business’s ability to manage data. But they can influence our approach towards managing that data as a valued enterprise asset.

The 7 Principles and how each of the 41 Standard Requirements aligns to them are shown below for convenience. However, it should be remembered that General Data as a function has many facets that interrelate closely together so there will be many dependencies between and across them. This dependency will also apply to the guidelines and so will be identified wherever appropriate.

The Standard focuses on the key things that we need to keep our data safe, secure and legal and what our external stakeholders would expect us to have in place.

Key Outcomes of implementing The Standard are:

* We know what data we need to run our business safely, securely and legally
* We know where that data is held
* We have clear ownership, accountabilities and governance for our data
* We have quality criteria established for our data, with ongoing monitoring and remediation
* We have the appropriate controls and assurance processes in place re: access / security / confidentiality / publishing etc.

## Principle 1 – Data is an Asset

Data (i.e. the representation of facts as text, numbers, graphics, images or sound or video) has a purpose, cost, value and lifecycle – it is an asset.

|  |  |
| --- | --- |
| **Data is an Asset** | |
| **Rationale**   * Our business is critically dependent on its data. It plays a key role in our operations, underpinning efficiency, productivity and robust decision making * Quality data directly drives our revenue and incentives * Data has a cost and hence a value – it is not just seen as an overhead nor simply a by-product of “IT systems” * It is important that the quality of our data is assured, communicated and maintained throughout its lifecycle * There needs to be clear accountability for data Modelling | **Implications**   * All forms of data (e.g. structured in databases, GIS, records, documents, photographs, historical data etc.) need to be managed as appropriate to meet our business needs and obligations * We need a formalised Modelling and control mechanism for data throughout its lifecycle, including risk Modelling * We need to record where we store and manage data, how it flows between systems, where it gets changed and how and when it is archived and / or disposed of * We need clear data owners and stewardship throughout its lifecycle with accountabilities and responsibilities clearly defined |
| **The Standard Requirements:**   * **1.1** Establish & maintain a relevant organisation structure with assigned data roles and responsibilities. * **1.2** Establish & maintain a General Data strategy & associated implementation plans including assurance, communications & engagement approaches * **1.3** Monitor and act on data ‘health’ issues via production of data quality metrics and delivery of action plans *[as appropriate]* * **1.4** Assess & manage business risks associated with data * **1.5** Establish, implement & maintain plans to protect data from loss, damage or unauthorised access * **1.6** Establish and maintain a data inventory | |

## Principle 2 – Data is Managed and Secure

All data is subject to ownership, governance and protection from unauthorised access throughout its full lifecycle (from planning, collection through to retention and disposal)

|  |  |
| --- | --- |
| **Data is Managed and Secure** | |
| **Rationale**   * The demand for reliable and accurate data is increasing and without the appropriate Modelling controls in place, the robustness of this data deteriorates with risks becoming less visible and uncontrolled * Effective data security requires us to manage both internal and external / unauthorised access * Data risk exposure needs to be identified and mitigated (e.g. security, quality, retention) | **Implications**   * We need to establish the appropriate governance structure and processes which have authority and control over the Modelling of data assets throughout their lifecycle, including disposal * We need to ensure we control access to our data, put measures in place to protect it from unauthorised access or tampering, and regularly test and review them * We need Modelling processes and the appropriate support to ensure compliance with internal and external obligations * We need to ensure we have the appropriate business continuity measures in place |
| **The Standard Requirements:**   * **2.1** Establish & maintain a General Data RACI covering the end to end data lifecycle & associated stakeholders * **2.2** Develop awareness of everyone’s data roles and responsibilities through regular communications and induction processes * **2.3** Categorise data for business criticality and confidentiality purposes * **2.4** Formalise and periodically review scope & effectiveness of data governance & assurance * **2.5** Restrict and monitor data access to authorised users only * **2.6** Record & investigate actual or suspected breaches of data security * **2.7** Dispose of data that is no longer required to be managed or retained (ensuring due consideration of any legal hold / retention requirements). * **2.8** Ensure that Vendors and Third-Party partners & suppliers comply with all National Grid General Data Standards | |

## Principle 3 – Data is Fit for Purpose

Data should be of the quality required for its intended uses

|  |  |
| --- | --- |
| **Data is Fit for Purpose** | |
| **Rationale**   * Data can only be regarded as fit for purpose if we know what the purposes of the data is and whether or not the data meets the required quality criteria for each and every purpose * We need to be mindful that data is collected for multiple uses across different processes and functions * It is important that any data we need to share / provide to our external stakeholders is robust and accurate – inaccurate or misreported data can lead to enforcement action and significant fines | **Implications**   * Data does not have to be “perfect” but must be fit for its intended purposes; thus the purposes need to be understood * In order to determine fitness for purpose we need to understand how and where the data is used * We need to define, monitor and address data quality metrics throughout the lifecycle of the data * We need to make it easy to report and correct inaccurate or missing data |
| **The Standard Requirements:**   * **3.1** Define the Data Owner for each area / set of data * **3.2** Define why the data is required, i.e. the business need for each area / set of data * **3.3** Define the quality criteria each area / set of data must meet * **3.4** Define who creates / modifies the data and in what process for each area / set of data * **3.5** Define how the quality is measured, validated, assured & improved for each area / set of data * **3.6** Define what the authoritative source is and how it is provided for each area / set of data * **3.7** Establish and maintain processes to report, record, identify root causes and correct inaccurate / inappropriate data | |

## Principle 4 – Data is standardised

Data should be standardised in terms of its definition, format, content and categorisation providing the ability to link differing forms of related data together

|  |  |
| --- | --- |
| **Data is standardised** | |
| **Rationale**   * Standard data definitions reduce the risk of misinterpretation and wasted analysis * Standard data definition and formats make it easier to link data from multiple sources * Standardised data definitions enable us to share and exchange data across our business processes and with external parties | **Implications**   * Data definitions are agreed * Data definition and standardisation rules are available and maintained * We need to understand the relationship between the differing data types across our business * We should have common system/data standards that allow use across multiple systems |
| **The Standard Requirements:**   * **4.1** Specify and maintain data definitions and requirements using agreed formats * **4.2** Establish and maintain a centralised repository of information about the data * **4.3** Establish and maintain a model identifying the relationships between the key data types *(i.e. through Data Models)* * **4.4** Establish and maintain a change control process for data definitions and models * **4.5** Verify that any new solution designs align with agreed definitions & models | |

## Principle 5 – Data has a single authoritative source

For all data there shall be a single and identified authoritative (master) source

|  |  |
| --- | --- |
| **Data has a single authoritative source** | |
| **Rationale**   * Across our organisation, differing groups, processes and systems need the same data and use it for multiple purposes. Data has to be considered a shared resource. * With a single trusted source of data in place we reduce the potential for conflicting data and the associated inefficiencies in resolving these differences * Versions of data that are not copied from a single master source increase the risk of conflicting outputs driving unnecessary reconciliation and re-work | **Implications**   * We need to understand what data we hold and where it is sourced from * We should avoid wherever possible making copies of data * Where we currently have multiple versions of similar data sets we need to determine which is the master source * We should only use copies of data that have a verifiable link to the agreed master source * Copies of data residing in other repositories should reflect the data in the authoritative source |
| The Standard Requirements:   * 5.1 Establish and maintain a catalogue of all available data sources * 5.2 Define the single (master) trusted source of data where there are multiple copies of a dataset * 5.3 Ensure that any replicated data used for reporting or analysis is traceable to the master source[s] | |

## Principle 6 – Data is Accessible

We should all have the appropriate access to the data we need to carry out our roles

|  |  |
| --- | --- |
| **Data is Accessible** | |
| **Rationale**   * In order to carry out our roles we all should be able to access the right data in the right format * We should be able to access critical data during prolonged outages and other business continuity events * It is important that access to data should only be limited to the data that is appropriate to our role / need | **Implications**   * We need to ensure we can access the appropriate data throughout its lifecycle with clearly defined service levels in place for outages etc. * We need to ensure we have the appropriate business continuity measures in place to access business critical data * We need to prevent access to data that is no longer required |
| **The Standard Requirements:**   * **6.1** Provide users with access to the data appropriate to their role [ensuring consideration of all legal & other compliance requirements] * **6.2** Establish & test business continuity & disaster recovery measures for Operationally Critical & Critical data * **6.3** Establish & maintain data ownership and documentation for archived data * **6.4** Define and maintain retrieval processes for archived data * **6.5** Provide a quality assessment for all business-critical data | |

## Principle 7 – Data is Published

Any data we publish should be defined, appropriate, quality assured and verifiable

|  |  |
| --- | --- |
| **Data is Published** | |
| **Rationale**   * We must publish and share data both internally and externally * We need to ensure we only publish data that is appropriate for its intended use / audience * There is an expectation that this data will be in the required format, and of sufficient quality * It is important we have a verifiable source for all data that we publish and share | **Implications**   * All published data should have a publishing owner * We need to implement clear governance around the release of data before it is published to ensure it meets the required standards (e.g. appropriate, accurate, complete and verifiable) * We need to ensure there are processes in place to manage any inaccuracies that are found in any published data * Information on the source, format, quality, owner etc. should be available for all published information |
| **The Standard Requirements:**   * **7.1** Assign a Publishing Owner for all published data * **7.2** Ensure all published data is appropriate for its intended use / audience * **7.3** Confirm the use of verifiable sources for all published data * **7.4** Validate the format & quality of all published data prior to release * **7.5** Establish & maintain any associated key information about the published data [including source, contacts, date of publication & any deviations from the agreed definitions] * **7.6** Establish & maintain appropriate control mechanisms for the Publishing Owner to approve release of external data publications * **7.7** Establish & maintain processes to report to the Publishing Owner any instances of inaccurate, incomplete or invalid published data with the appropriate actions taken | |

National Grid Document Taxonomy

It is important to know which National Grid policies and other standard documentation should be applied to Data Modelling. Therefore, employees within the business should be able to easily find and navigate to the appropriate policies and procedures according to their role. To utilise this efficiently it would be useful to have some standardisation. The definitions below are those which tend to be widely used across many other businesses. Particularly those subject to regulation and HSE scrutiny like ours.

**Standards** such as the BMS represent the goal for an organisation, such as the rules that we should adhere to.  They set out what must be done and ensure there is clarity over what is required. They help identify the things that really mater – the areas of highest value and/or highest risk. The Standard defines the bar against which progress will be measured not simply ‘what can be achieved today’. It provides a platform for sharing best practice and creates a foundation for effective assurance across the business.  There will often be a road map to achieving a Standard. Sometimes a Standard can be external yet will still set out this target: such as an ISO or British Standard. A Standard can have a number of components such as Principles and Requirements.

Our Standards are split into 3 categories: **Capability**, **Functional** and **Risk-based** standards.

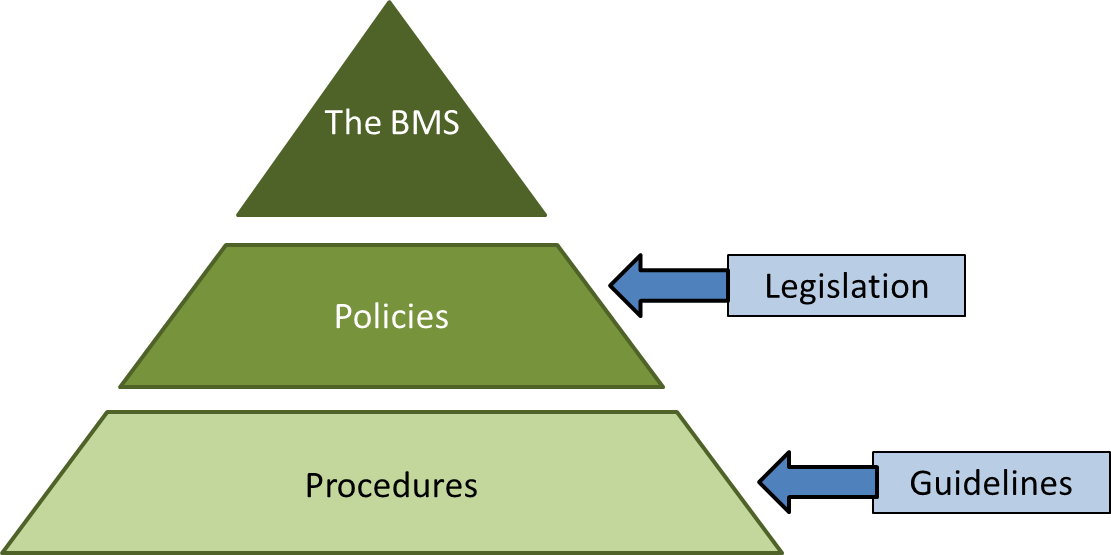
The General Data Standard under the Capability category and is primarily focused on building our capability as a business by learning from each other and sharing best practice. For more information on standards, please see The National Grid Book accessible from the main Grid Home Page.

**Policies** cover what the business has decided to do to meet a standard. Policies detail the “must do” items. The policy should reflect what is achievable at the same time as supporting the delivery of the Standard. There should be traceability back to the Standard. Whilst achieving the Standard may be a journey, we must always be compliant to the policies. The policies are normally set at a Business Entity level. As well as driving compliance to the BMS, the policies will also reflect any immoveable, external requirements such as HSE legislation or regulatory requirements. A policy should clearly state who it applies to. A policy may have several requirements (sometimes historically known as “standards”) that the business should comply with in order to meet the policy.

**Procedures** should provide the steps to be followed that will result in the delivery of/compliance to, the policies. There should be traceability back to the Standard. Procedures are normally written at a more granular level. They may be owned by sub-sections of an entity. Procedures should clearly state roles and responsibilities of people and departments.

**Guidelines** are reference documents containing tools that can be used to support the delivery of the Procedures and policies.

The diagram below shows the possible relationship between these taxonomies.



Document Administration

|  |  |
| --- | --- |
| Owner | Graham South |
| Author/s first edition | Gavin Marshall |
| Date first created | 11 January 2022 |
| Date of current edition | 11 January 2022 |
| Date last amended | 11 January 2022 |

Reviewers

|  |  |  |
| --- | --- | --- |
| Review Name | Role | Date & version |
|  |  |  |
| Ken Jones | Head of Data Governance | 11 January 2022 v1.0 |
| Kam Bharj | Data Governance Senior Manager | 11 January 2022 v1.0 |
| ??? | ?? Group Head of Data & Analytics | 11 January 2022 v1.0 |
|  |  |  |

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

1. Quote from the Data Management BMS [↑](#footnote-ref-2)