Your Online Data Architecture

Metadata and Supporting Information Requirement

Richard Dobson

Practice Manager Data Systems

25/06/2020

Foreword

The path to a zero-carbon future is exciting, challenging and realising this ambition will call for innovative and disruptive ideas, alongside maximising proven sustainable and low carbon solutions. To ensure the UK achieves our carbon neutral commitments, it is critical that the UK energy sector maximises the value of existing (and future) national infrastructure, resources and the significant levels of existing and emerging data.

Realising this value however will call upon the energy sector to undertake a data-enabled cultural evolution. An evolution which actively enables open-data exchange, utilising data fed from multiple energy vectors, encompassing the breadth of the energy system and the plethora of stakeholders within. A true 'whole systems' approach which will give rise to a secure and managed service, a service affording access to the central data exchange for stakeholders wishing to illicit positive, innovative and lasting change within the energy sector.

Siemens, together with its partners the Energy Systems Catapult (ESC) and National Innovation Centre for Data (NICD) will deliver a 'digitally integrated energy system' which supports a Common Data Architecture concept. Underpinning the vision is an implementation of the open-data platform, constructed upon the requirements of the users and employing a sector specific metadata standard to drive commonality, enabling data-exchange. To do this Siemens will create 'Your Online Digital Architecture'.

The platform will be constructed upon the three relevant building blocks identified within the report 'Energy Data Taskforce: A Strategy for a Modern Digitalised Energy System' – incorporating asset registration strategy, data catalogue and digital system mapping.

Siemens promotes an inclusive approach to successful deployment, one which will be employed from beginning to end, facilitating user requirement capture workshops alongside show and tell events to provide insight toward the project outcomes and providing best practice guidelines. These events will support organisations who wish to utilise and embed the insight and outcomes, covering topics such as data transparency, data licensing and liability wavers, and data obfuscation / data protection techniques. Each of which are reflective of the metadata standard and Common Data Architecture underpinning 'Your Online Digital Architecture'.

The challenge is clear – the true value in data, in support of the transition to a low carbon economy, is in enabling visibility, access and insight throughout the energy value-chain – the industry must embrace this opportunity as a true catalyst for change, creating an open, yet secure, data marketplace which will create a modern, digitalized, energy system – one which drives system cost reduction, increases levels of asset visibility, improves system stability, provides capability for informed system management approaches and enables impactful innovation at scale. All critical factors in a decarbonized, digitalized and decentralized energy system.

Andrew Smyth - Head of Customer Success, Software

Siemens Smart Infrastructure

Introduction

After carrying out stakeholder workshops, it has become clear that there is a need for data which is made readily available, to be well documented and accompanied by all relevant information to make the data usable. In addition, the data best practice guidance which was developed as part of the Modernising Energy Data programme, has specifically called out Metadata and Supporting information as key parts of Data Best Practice.

In the following report, we will discuss the investigation which has been conducted during this discovery phase and make some clear recommendations for challenges which must be addressed in the Alpha project.

Metadata

The Data Best Practice guidance sets out the importance of Metadata for the energy sector.

> 3. Describe data accurately using **industry standard metadata**
>
> To realise the maximum value creation from data within an organisation, across an industry or across the economy actors need to be able to understand basic information that describes each dataset. To make this information accessible, the descriptive information should be structured in an accepted format and it should be possible to make that descriptive information available independently from the underlying dataset.
>
> **Metadata** is a dataset that describes and gives information about another dataset.

The YODA platform will rely on metadata to provide descriptions of datasets which are registered on the platform and also aid the linking of related datasets which have not yet been mapped onto the underlying metamodel.

For this we propose to initially use the Dublin Core 'Core Elements' metadata standard (Dublin Core) ISO 15836-1:2017 as recommended by the Energy Data Taskforce and the Data Best Practice Guidance. We believe this balances deliverability against functionality and provides a strong foundation to build upon.

There are 15 'core elements' as part of the Dublin Core standard which are described as follows:

Element	Description
Title	Name given to the resource
Creator	Entity primarily responsible for making the resource
Subject	Topic of the resource *(e.g. Keywords from an agreed vocabulary)*
Description	Account of the resource
Publisher	Entity responsible for making the resource available
Contributor	Entity responsible for making contributions to the resource
Date	Point or period of time associated with an event in the lifecycle of the resource
Type	Nature or genre of the resource

Element	Description
Format	File format, physical medium, or dimensions of the resource
Identifier	Compact sequence of characters that establishes the identity of a resource, institution or person alone or in combination with other elements *e.g. Uniform Resource Identifier (URI) or Digital Object Identifier (DOI)*
Source	Related resource from which the described resource is derived *(e.g. Source URI or DOI)*
Language	Language of the resource *(Selected language(s) from a agreed vocabulary e.g. ISO 639-2 or ISO 639-3).*
Relation	Related Resource *(e.g. related item URI or DOI)*
Coverage	Spatial or temporal topic of the resource, spatial applicability of the resource, or jurisdiction under which the resource is relevant
Rights	Information about rights held in and over the resource

Core descriptions from Dublin Core Metadata Initiative (DCMI) licenced via CC BY 3.0 - edits or additions are made in *italics*.

Within the Data Best Practice guidance some notes have been provided which guide the implementation of the standard in the energy sector. We do not repeat this guidance here but build upon it in the context of a common data architecture.

Element	Common Data Architecture Considerations
Creator, Publisher, Contributor	• The CDA should include an **authoritative list of organisations** (past and present) and provide the option to register new organisations. Where an organisation is a registered participant in the CDA solution this should link to their account. • This enables prospective data users to understand the provenance of the datasets and gives a route of escalation for queries.
Subject	• The CDA provider should develop and promote the use of a **common glossary of terms** in line with the DFTG recommendation for a Reference Data Library and the findings of the Data Visibility Project being led by the ONS. • This needs to be an open and collaborative project which draws on existing standards (such as CIM), legal definitions (e.g. licences) and informal definitions used across industry. Only through a collaborative project will this gain the traction required to drive standardisation • This is the first stage towards standardisation of data across the energy sector and enables data to be searched and linked more effectively.

Element	Common Data Architecture Considerations
Format	• The format should refer to the format of the delivered data rather than the access method e.g. if an API request returns a csv then the format should refer to the csv.
Unique Resource Identifier (URI)	• This should be a globally unique reference within the CDA system.
Source	• Where possible, the source field should be refer to the URI of the resource (or resources) within the CDA system, this will enable the provenance of data to be tracked and enable users to choose if they would rather work with the more raw or processed version of a dataset
Relation	• The relation field should be used to link to related resources such as supporting information (e.g. data collection methodology, reference data, processing scripts, etc.) which may or may not be made available through the common data architecture solution. • Where resources are available through the common data architecture solution these should be referenced using their URI. • The CDA solution should enable common reference datasets (e.g. data standards, unit conversion tables, etc.) to be given a CDA URI even if they are not accessible through the system
Rights	• As discussed in section **Error! Reference source not found.**, the CDA should offer a set of standard licences which can be used by data providers

As recommended by the Data Best Practice Guidance, metadata should be stored in an independent file from the original data and in a machine readable format, such as JSON, YAML or XML that can easily be presented in a human readable format using free text editors. This approach ensures that metadata can be shared independently from the dataset, that it is commonly accessible and not restricted by software compatibility.

Supporting Information

The Data Best Practice guidance sets out the need for supporting information to accompany datasets.

> 4. Enable potential users to understand the data by providing **supporting information**
>
> When data is published openly, made publicly available or shared with a specific group it is critical that the data has any supporting information that is required to make the data useful for potential users. There is a need to differentiate between **Core Supporting Information**, without which the data could not be understood by anyone, and **Additional Supporting Information** that makes understanding the data easier. As a rule of thumb, if the original custodian of the dataset were to stop working with it and then come back 10 years later with the same level of domain expertise, but without the advantage of having worked with the data on a regular basis, the Core Supporting Information is that which they need to make the dataset intelligible.

Data Custodians should make Core Supporting Information available with the dataset.

It is vital that supporting information is provided by data custodians alongside published data as this is key to enable potential data users to understand the information contained within. This should be referenced within the relation metadata field.

Summary

In order to achieve a successful implementation of the YODA platform, it is important that all datasets and information shared is accompanied with any metadata and supporting information in line with the Energy Data Taskforce best practice report. The Energy Data Taskforce recommends that the Dublin Core 'Core Elements' should be implemented initially to gain a metadata standard and the metadata should be stored in an independent file, presented in a human readable format.

In the movement to publicly open data, it is critical that supporting information is accessible with Core Supporting Information and Additional Supporting Information being provided where necessary. Core supporting information must be made readily available by the data custodians and should be referenced within the relation metadata field.

Energy Systems Catapult supports innovators in unleashing opportunities from the transition to a clean, intelligent energy system.

Energy Systems Catapult
7th Floor, Cannon House
18 Priory Queensway
Birmingham
B4 6BS

es.catapult.org.uk
info@es.catapult.org.uk
+44 (0)121 203 3700