Making a positive difference for energy consumers

Guidance

Data Best Practice Guidance

Publication date:	7 August 2023	**Contact:**	Charley Clark, Senior Manager
		Team:	Energy System Data Regulation
		Tel:	
		Email:	Digitalisation@ofgem.gov.uk

This document is Ofgem's Data Best Practice Guidance. Its requirements must be complied with by companies whose licences are subject to the RIIO-GD2, RIIO-T2, and RIIO-ED2 price controls. It is part of Ofgem's standards for data and digitalisation, to be followed by relevant Licensees and by Ofgem itself on a voluntary basis.

Document version	Description	Changes since previous document
Data Best Practice Guidance v0.3	Proposed version May 2021	N/A
Data Best Practice Guidance v1	Version as a result of consultation May 2021	As set out in "Track Changes DBP Guidance v1.00"
Track Changes Data Best Practice Guidance v1.1 for consultation	Proposed version February 2023	Edits made to ensure Electricity Distribution licensees are required to follow this guidance document. Edits made to principles 3, 5, 9, and 11. Intended outcomes introduced for each Principle.
Track Changes Data Best Practice Guidance v2.0 for decision	Version as a result of consultation February 2023	As set out in "Track Changes Data Best Practice Guidance v1.1 for consultation"

Contents

1. Introduction

1.1. This document provides principles and explanations that describe requirements for complying with Data Best Practice Guidance ("DBP Guidance").

1.2. This document is directed at gas distribution network companies, gas and electricity transmission network companies, electricity distribution network companies, and the electricity system operator (for the purposes of this document 'network companies' or 'licensees') as well as their stakeholders. The licensees are required to comply with DBP Guidance when working with, or making decisions about, Energy System Data[1].

1.3. To find out more about this licence obligation, please visit these following links:

- RIIO-2 Final Determinations for Transmission, Gas Distribution and Electricity System Operator[2];
- RIIO-ED2 Final Determinations for Electricity Distribution licensees[3];
- Decision on the proposed modifications to the RIIO-2 Transmission, Gas Distribution and Electricity System Operator licences[4].; and
- Decision on the proposed modifications to the RIIO-2 Electricity Distribution licences[5].

1.4. DBP Guidance is part of Ofgem's standards for data and digitalisation, to be followed by relevant licensees and by Ofgem itself[6]. Both the Digitalisation Strategy and Action Plan Guidance and DBP Guidance are part of our standards for data and digitalisation.

1.5. The work of the other organisations, such as GO FAIR and Government Digital Service (GDS) strongly informed the initial development of DBP Guidance. The GDS provides wide-ranging support for topics relating to data and digitalisation; it gives information

[1] Special Licence Condition 9.5 of the Gas Distribution, Gas Transmission, Electricity Transmission, and Electricity Distribution licences, and Special Licence Condition 2.11 of the Electricity System Operator licence
[2] https://www.ofgem.gov.uk/publications-and-updates/riio-2-final-determinations-transmission-and-gas-distribution-network-companies-and-electricity-system-operator
[3] https://www.ofgem.gov.uk/publications/riio-ed2-final-determinations
[4] https://www.ofgem.gov.uk/publications-and-updates/decision-proposed-modifications-riio-2-transmission-gas-distribution-and-electricity-system-operator-licences
[5] https://www.ofgem.gov.uk/publications/decision-proposed-modifications-riio-2-electricity-distribution-licences
[6] https://www.ofgem.gov.uk/publications-and-updates/forward-work-programme-202122.

and methods that span all the principles in DBP Guidance. The following are particularly relevant:

- GO FAIR and its FAIR data principles[7];
- GDS Service Standard[8];
- GDS Technology Code of Practice9; and the
- GDS Service Manual[10].

[7] https://www.go-fair.org/fair-principles/
[8] https://www.gov.uk/service-manual/service-standard
[9] https://www.gov.uk/government/publications/technology-code-of-practice/technology-code-of-practice
[10] https://www.gov.uk/service-manual

2. Summary

2.1. DBP Guidance is designed to ensure data is treated as an asset and used effectively for the benefit of consumers, stakeholders and the Public Interest. It is a principles-based approach which provides guidance on the quality, accuracy and accessibility of data. It includes the principle that Data Assets must be treated as Presumed Open[11] which means that data must be made available for all people to use, unless the licensee responsible for handling the data provides specific evidence to show that the data should be withheld or its availability reduced (e.g. to protect individuals' rights to privacy). By complying with DBP Guidance licensees will enable the full benefits of data to be unlocked for consumers.

Data Best Practice principles

1. Identify the roles of stakeholders of Data Assets.
2. Use common terms within Data Assets, Metadata and supporting information.
3. Describe data accurately using industry standard Metadata.
4. Enable potential Data Users to understand Data Assets by providing supporting information.
5. Make Data Assets discoverable for potential Data Users.
6. Learn and deliver to the needs of current and prospective Data Users.
7. Ensure data quality maintenance and improvement is prioritised by Data User needs.
8. Ensure Data Assets are interoperable with Data Assets from other data and digital services.
9. Protect Data Assets and systems in accordance with Security, Privacy and Resilience (SPaR) best practice.
10. Store, archive and provide access to Data Assets in ways that ensure sustained benefits.
11. Treat all Data Assets, their associated Metadata and Software Scripts used to process Data Assets as Presumed Open.

[11] https://es.catapult.org.uk/reports/energy-data-taskforce-report/

Definitions

Creative Commons Attribution Licence[12]: A data sharing licence that allows Data Users to access, share, and adapt a Data Asset as long as an appropriate reference is made to the Data Custodian.

Data Asset: Any entity that is comprised of data. For example, a database is a data asset that is comprised of data records. A data asset may be a system or application output file, database, document, or web page. A data asset also includes a service that may be provided to access data from an application. For example, a service that returns individual records from a database would be a data asset. Similarly, a web site that returns data in response to specific queries (e.g., www.weather.com) would be a data asset.

This definition is taken from National Institute of Standards and Technology (NIST).[13]

Data Catalogue: An informative and searchable inventory of all Data Assets for which the licensee is the Data Custodian[14]. This allows users to search and identify key Metadata associated with a Data Asset. It should be kept live, updated, and be an accurate reflection of the licensee's Data Assets.

Data Contact Point: An organisation or individual who is the primary point of contact about a Data Asset or Metadata associated with a Data Asset.

Data Controller: A person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of a specific Data Asset.

This is based on the Information Commissioner's Office (ICO) definition but has been modified by removing reference to personal data and replacing it noting the processing of a Data Asset.[15]

Data Custodian: A person, public authority, agency or other body that has a legal right to process and publish a Data Asset as the Data Controller or otherwise.

Data Processor: The natural or legal person, public authority, agency or other body which processes Data Assets on behalf of the Data Controller.

---

[12] https://creativecommons.org/licenses/by/4.0/
[13] https://csrc.nist.gov/glossary/term/data_asset
[14] Adaptation of the IBM definition for a Data Catalogue - https://www.ibm.com/uk-en/topics/data-catalog
[15] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/controllers-and-processors/what-are-controllers-and-processors/

This is based on the ICO definition but has been modified by removing reference to personal data and replacing it noting Data Assets.[16]

Data Subject: The identified or identifiable living individual or entity to whom data relates.

Data User: An organisation or individual which utilises data held by a Data Custodian for any reason.

Data Best Practice Guidance: means (1) the guidance document issued by the Authority[17] in accordance with Part D of; Special Condition 9.5 (Digitalisation) of the RIIO-2 price controls for Electricity Transmission, Gas Transmission and Gas Distribution, Special Condition 2.11 (Digitalisation) of the RIIO-2 price controls for the Electricity System Operator, and Special Condition 9.5 (Digitalisation) of the RIIO-ED2 price control for Electricity Distribution and (2) part of Ofgem's standards for data and digitalisation.

Digitalisation Action Plan: a licensee's plan to digitalise its Products and Services prepared and published in accordance with Part B of; Special Condition 9.5 (Digitalisation) of the RIIO-2 price controls for Electricity Transmission, Gas Transmission and Gas Distribution, Special Condition 2.11 (Digitalisation) of the RIIO-2 price controls for the Electricity System Operator, and Special Condition 9.5 (Digitalisation) of the RIIO-ED2 price control for Electricity Distribution.

Digitalisation Strategy: the strategic approach taken by a licensee to digitalise its Products and Services and evidenced by the archive prepared and published by the licensee in accordance with Part A of; Special Condition 9.5 (Digitalisation) of the RIIO-2 price controls for Electricity Transmission, Gas Transmission and Gas Distribution, Special Condition 2.11 (Digitalisation) of the RIIO-2 price controls for the Electricity System Operator, and Special Condition 9.5 (Digitalisation) of the RIIO-ED2 price control for Electricity Distribution..

Digitalisation Strategy and Action Plan Guidance: means (1) the guidance document issued by the Authority in accordance with Part C of; Special Condition 9.5 (Digitalisation) of the RIIO-2 price controls for Electricity Transmission, Gas Transmission and Gas Distribution, Special Condition 2.11 (Digitalisation) of the RIIO-2 price controls for the Electricity System

[16] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/controllers-and-processors/what-are-controllers-and-processors/
[17] The terms "the Authority", "we", and "us" are used interchangeably in this document

Operator, Special Condition 9.5 (Digitalisation) of the RIIO-ED2 price control for Electricity Distribution and (2) part of Ofgem's standards for data and digitalisation.

DSAP: A combination of both Digitalisation Strategy and Digitalisation Action Plan.

Dublin Core Metadata Standard: A standard for formatting and structuring Metadata. Information on best practice with regard to implementation of Dublin Core[18] can be found in the Data Best Practice supporting information document.

Energy System Data: all Data Assets for which an entity is a Data Custodian as a consequence of it exercising its rights and obligations under a licence granted under section 6 (1) or (1A) of the Electricity Act 1989 or section 7, 7ZA, 7A or 7AB of the Gas Act 1986".

Metadata: a set of data that describes and gives information about other data.

Open Data: Data Assets, their associated Metadata and Software Scripts used to process Data Assets that are made available for anyone to use, modify and distribute without restrictions.

Open Data Triage: The process carried out by a Data Custodian to determine if there is any evidence of sensitivities associated with Data Assets, their associated Metadata and Software Scripts used to process Data Assets if they are used as Open Data. These sensitivities are limited to those that:

(i). relate to peoples' rights to personal privacy;
(ii). security needs;
(iii). obligations from legislation and/or regulation;
(iv). commercial requirements that, if not protected, will have a negative impact on Products and Services for end-consumers; and
(v). would have a negative impact on the Public Interest.

Where any of the above sensitivities are identified, Open Data Triage will also include the determination of how the Data Custodian can mitigate any risk associated with them, while also making the Data Assets, their associated Metadata and Software Scripts used to process

[18] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

Data Assets as open to stakeholders as possible. The Data Custodian should consider both processing of and/or whether providing different levels of access by different types of stakeholders to the Data Assets, their associated Metadata and Software Scripts used to process Data Assets would help to mitigate any identified risk.

Open Government Licence[19]: A data sharing licence that allows Data Users to access, share, and adapt a Data Asset as long as appropriate reference is made to the Data Custodian.

Presumed Open: The treatment of Data Assets, their associated Metadata and Software Scripts used to process Data Assets as Open Data, subject to Open Data Triage.

Products and Services: Anything that a licensee can offer to a market for attention, acquisition, use or consumption that could satisfy a need or want.

Public Interest: The welfare or well-being of the general public and society.

Single Provider Product or Service: A product or service among the Products and Services provided by a licensee where no alternative option or provider is available to parties seeking to access that product or service.

Software Scripts: A code and its programming documentation; including information on how to execute that code, that enables Data Users to read, capture, process, store or transmit a Data Asset or Metadata.

the Authority: means the Gas and Electricity Markets Authority that is established under section 1 of the Utilities Act 2000

[19] https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

3. The Data Best Practice Principles

1. Identify the roles of stakeholders of Data Assets

Explanation

3.1. The licensee must identify the Data Assets that it is the Data Custodian of; for these, the licensee must also identify any relevant Data Subjects, Data Controllers and Data Processors. The licensee must keep this information in logs.

Intended outcome

3.2. The licensee has a process for identifying the roles in its organisation relating to Data Assets and can demonstrate how this process is applied to all Data Assets for which the licensee is the Data Custodian.

3.3. The licensee has a log(s) identifying its Data Assets using the ontology from 3.2.

2. Use common terms within Data Assets, Metadata and supporting information

Explanation

3.4. The licensee must enable Data Users to search for and link Data Assets and associated Metadata to Data Assets and Metadata provided by other licensees. The licensee must label and describe Data Assets and Metadata using a taxonomy that is commonly recognised by practitioners who use the Metadata across the relevant subject matter domain.

Intended outcome

3.5. The licensee has a taxonomy for describing Data Assets, Metadata, and supporting information.

3.6. The licensee can demonstrate how this taxonomy is "commonly recognised" by practitioners, either through industry engagement or commonality between taxonomy use across the sector.

3. Describe data accurately using industry standard Metadata

Explanation

3.7. The licensee must make it easy for Data Users to be able to use and understand information that describes each Data Asset. The licensee must therefore provide Metadata associated with Data Assets and this Metadata must be made available to Data Users independent of the Data Asset. This Metadata must be made accessible through a Data Catalogue.

3.8. The licensee must treat the Metadata as a Data Asset. When providing Metadata, the licensee must format and structure this Metadata using the latest version, or a subsequent iteration, of the Dublin Core Metadata Standard.

3.9. There is no requirement for the licensee to create Metadata about its Metadata associated with Data Assets.

3.10. When it updates or extends a Data Asset, the licensee must ensure that the Metadata reflects any such changes so that Data Users can identify additions or changes.

Intended outcome

3.11. The licensee has a Data Catalogue in a location which is accessible to current and prospective Data Users. The Data Catalogue utilises the Dublin Core Metadata Standard.

3.12. The licensee can demonstrate how it uses the latest version, or a subsequent iteration, of the Dublin Core Metadata Standard when utilising and providing Metadata.

3.13. The licensee can demonstrate its process for updating Metadata in response to an update or extension of a Data Asset.

4. Enable potential Data Users to understand the Data Assets by providing supporting information

Explanation

3.14. Throughout the lifecycle of a Data Asset the licensee must make available supporting information that Data Users require for the benefits to be gained by consumers, stakeholders and the Public Interest. The licensee must ensure a point of contact is provided for Data Users to raise and resolve enquiries about the Data Asset and its supporting information.

Intended outcome

3.15. The licensee can demonstrate that they provide clear and concise supporting information to both prospective and current Data Users, alongside their Data Assets.

3.16. The licensee has a point of contact that assists with enquiries relating to a Data Asset and its supporting information, and can demonstrate how feedback provided to this point of contact helps ensure its supporting information meets the needs of Data Users.

5. Make Data Assets discoverable to potential Data Users

Explanation

3.17. The licensee must ensure that any potential Data Users can identify the Data Assets that the licensee is the Data Custodian of, and how Data Users can pursue access to these Data Assets. The licensee must ensure that the Metadata associated to Data Assets is discoverable to Data Users, subject to the outcome of an Open Data Triage process. Data Assets should be made discoverable through a Data Catalogue.

Intended outcome

3.18. The licensee has a Data Catalogue in a location accessible to Data Users. The Data Catalogue states access levels for different types of Data User.

3.19. The licensee publishes Metadata through their Data Catalogue allowing users to identify the contents of Data Assets.

6. Learn and deliver to the needs of current and prospective Data Users

Explanation

3.9. The licensee must identify the Product and Service requirements of Data Users who use, or who wish to use, the Data Assets provided by it as Data Custodian. The licensee must then develop and deliver Products and Services to meet Data Users' requirements, where doing so is of reasonable cost and would benefit at least one of the following: end-consumers, stakeholders, or the Public Interest. Where the licensee is obligated to publish and update a DSAP, these Products and Services must be integrated into those publications.

Intended outcome

3.10. The licensee can demonstrate its process for gathering information on user needs from its Data Users.

3.11. The licensee can demonstrate its process for assessing what Products and Services are needed to meet the needs of Data Users, and whether these Products and Services deliver benefits for end-consumers, stakeholders or the Public Interest.

7. Ensure data quality maintenance and improvement is prioritised by Data User needs

Explanation

3.12. Where the licensee reasonably expects the Data Users' application of Data Assets for which it is the Data Custodian to deliver a net benefit for end-consumers, stakeholders and/or the Public Interest, the licensee must ensure that Data Assets are of a quality that is sufficient to meet reasonable requirements of its Data Users. Data Users must have an option for contesting decisions regarding the definition of sufficient data quality of a Data Asset.

3.13. Where data quality issues are identified the licensee must ensure that these issues are logged, considered and rectified as soon as practicable.

Intended outcome

3.14. The licensee can demonstrate it has processes in place for assessing, and ensuring, data quality and can demonstrate that these processes meet the requirements of its Data Users.

3.15. The licensee has a log of data quality issues detailing how these issues were, or will be, resolved.

8. Ensure Data Assets are interoperable with Data Assets from other data and digital services

Explanation

3.16. The licensee must enable interoperability, between the Data Assets for which it is Data Custodian and Data Assets of other licensees as a minimum standard.

3.17. When the licensee makes Data Assets available, it must do so in ways that make it reasonably easy for Data Users to gain information and/or insight from those Data Assets in conjunction with Data Assets from other licensees. There must also be sufficient information to align to Data Assets from other industries.

3.18. The licensee must make data available in such a way that it is reasonably easy for Data Users to:

(i). exchange Data Assets between systems;
(ii). interface with Data Assets held in the licensee's systems; and
(iii). join Data Assets with other Data Assets, such as by using standard interfaces, standard data structures and/or common reference data.

Intended outcome

3.19. The licensee can demonstrate how they have created, and will be creating, the Products and Services necessary to enable interoperability.

3.20. The licensee can demonstrate how they have built their Products and Services so that Data Users can easily join their Data Assets with Data Assets of other licensees.

9. Protect Data Assets and systems in accordance with Security, Privacy and Resilience best practice

Explanation

3.22. The licensee must ensure adherence to relevant regulations and legislation in relation to cyber security and SPaR requirements.

Intended outcome

3.23. The licensee can demonstrate how its Products and Services are developed in compliance with the current regulations and legislation relating to cyber and physical security.

10. Store, archive and provide access to Data Assets in ways that ensures sustained benefits

Explanation

3.24. When Data Assets are not required by the licensee, the licensee must ask stakeholders whether they consider that the Data Assets could create a future benefit if archived. The licensee must archive Data Assets when, taking account of stakeholders' views it determines that the storage will be a net benefit to consumers, stakeholders and/or the Public Interest.

3.25. When archiving, the licensee must also ask stakeholders for views on the storage method and formats to use. In determining what to archive, it must consider:

- Data Assets;
- Metadata;
- Software Scripts used to process Data Assets;
- data derived resulting from this processing of the original Data Asset; and
- human-readable representations of the data and any other relevant supporting information.

3.26. The licensee must ensure that the risk of unintentional or malicious deletion of Data Assets, Metadata and Software Scripts used to process Data Assets is effectively managed and monitored to ensure possible recovery.

Intended outcome

3.27. The licensee can demonstrate its process for archiving Data Assets and how this process has been built in consultation with Data Users.

3.28. The licensee can demonstrate regular engagement with Data Users on the archival of Data Assets.

11. Treat all Data Assets, their associated Metadata and Software Scripts used to process Data Assets as Presumed Open

Explanation

3.29. The licensee must treat all Data Assets, their associated Metadata and Software Scripts used to process Data Assets where it is the Data Custodian, as Presumed Open and these must be subjected to Open Data Triage.

3.30. The licensee must treat information created during Open Data Triage as Open Data, except where this will result in a sensitivity listed in the Open Data Triage definition.

3.31. Where a sensitivity is identified with the Data Assets, their associated Metadata and Software Scripts used to process Data Assets, the licensee must take all reasonable steps to provide suitable options to make them available in a format or version that mitigates the risk associated with any identified sensitivity. When identifying those options the licensee should additionally consider whether providing different stakeholders with different levels of access would mitigate any identified risk while minimising any reduction in the utility of the Data Asset.

3.32. The licensee must make available the Data Assets, their associated Metadata and Software Scripts used to process Data Assets in the changed formats, versions or with the different levels of access to stakeholders, where it is beneficial to end consumers, stakeholders and the Public Interest to do so.

3.33. The licensee when making Open Data available to current and prospective Data Users, must do so utilising the latest version, or a subsequent iteration, of the Creative Commons Attribution Licence or the Open Government Licence.

3.34. The licensee must record at least the following information about Open Data Triage processes:

- what has been triaged;
- when the process took place;
- a description of the sensitivities and risks, if any, that have been identified including the type of sensitivity as defined by Open Data Triage;
- the options considered for how to mitigate any sensitivities or risks identified and the impact these have on the utility of the Data Assets, their associated Metadata and/or Software Scripts used to process Data Assets; and

- any decisions made.

3.35. The licensee must ensure there is a point of contact available to stakeholders to allow them to seek information about Open Data Triage processes as well as to provide them with the opportunity to challenge decisions and escalate issues.

3.36. The licensee must keep under review its collection of available Data Assets, their associated Metadata and/or Software Scripts used to process Data Assets for risks or sensitivities and must mitigate these as they arise.

Intended outcome

3.37. The licensee can demonstrate that it applies the Open Data Triage process as defined in DBP Guidance.

3.38. The licensee can demonstrate that it makes Data Assets available as Open Data by default and takes all reasonable steps to provide suitable options to make Data Assets available in a format or version that mitigates the risk associated with any identified sensitivity.

3.39. The licensee can demonstrate that it shares Open Data using the latest version, or a subsequent iteration, of the Creative Commons Attribution Licence or the Open Government Licence.

3.40. The licensee has a point of contact that assists with enquiries relating to a Open Data Triage.