Innovate UK

Net Zero Living Data Guidance

16/03/2023

Contents

Glossary

Creative Commons Attribution Licence: 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).

Data Best Practice: The Office for Gas and Electricity Market (Ofgem)'s Data Best Practice Guidance.

Data Catalogue: An informative and searchable inventory of all Data Assets for which the organisation is the Data Custodian. 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 organisation's Data Assets.

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.

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 Ethics: A branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use.

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

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.

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

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

Net Zero: A state in which the total emissions of greenhouse gases are equal to or less than those we remove from the environment.

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 relate to:

(i) peoples' rights to personal privacy;
(ii) security needs;
(iii) obligations from legislation and/or regulation;
(iv) commercial requirements or intellectual property that, if not protected, will have a negative impact on end-consumers;
(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 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: 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.

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 Guidance: The Innovate UK Net Zero Living Data Guidance.

1. Context

This Guidance is created to improve data maturity among innovators funded by the Net Zero Living programme of Innovate UK. Its objective is to enable organisations across the UK to benefit from data in the Net Zero domain and capitalise on the relevant data-driven economic opportunities by unlocking the economic and societal benefits of data while protecting our values.

It builds on the commitment Innovate UK made in the Government's Energy Digitalisation Strategy to 'give greater visibility of learning and findings from funded energy innovation projects, including embedding Energy Data Best Practice principles'. To facilitate this, a series of principles, derived from the Office for Gas and Electricity Market's (Ofgem) Data Best Practice principles, have been adopted for the purposes of The Guidance. The principles of The Guidance are set out in section three. These principles are similar to, but not exactly the same as, Ofgem's Data Best Practice principles.

The Guidance also supports retaining appropriate data ownership, data security and intellectual property rights of organisations providing data. In addition, The Guidance and the associated material have been curated to support organisations in developing stronger data and digital capabilities and are intended to accelerate capability building of said organisations or projects while avoiding onerous reporting requirements.

The principles are provided with guidance, tools and additional support and are available to those who are required to adhere to The Guidance. This assortment of support is set out in section four.

1.1   Scope of The Guidance

The Guidance applies to organisations which gather, create or update Data Assets, Metadata, and Software Scripts relating to project activity as a result of contractual arrangements with Innovate UK covered under the Net Zero Living programme of awards.

1.2   Wider Government Objective Alignment

The objectives of The Guidance align with the UK's National Data Strategy, which is the government's framework strategy to drive the UK in building a world-leading data economy while ensuring public trust in data use. Specifically, The Guidance addresses mission one of the Strategy, which is to unlock the value of data through making it usable, accessible and available across the economy.

The Guidance can be read in conjunction with UKRI's Guidance on Best Practice in the Management of Research Data, which sets out UKRI's expectations on award holders of research projects.

2. Adherence to The Guidance

Where organisations have been instructed to adhere to The Guidance through contractual arrangements with Innovate UK, it is expected that those organisations employ reasonable endeavours to comply with the principles of The Guidance.

Where funding has been received, it is expected that all data gathered, created or updated relating to project activity will be triaged through an Open Data Triage and made openly available on a publicly accessible website where possible, taking into account the sensitivities listed in the Open Data Triage definition.

3. Principles

1. Identify the roles of stakeholders of Data Assets.

Explanation

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

Intended outcome

(b) The organisation 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 organisation is the Data Custodian.

(c) The organisation has a log(s) identifying its Data Assets using the ontology from 1(b).

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

Explanation

(a) The organisation must enable Data Users to search for and link Data Assets and associated Metadata to Data Assets and Metadata provided by other organisations. The organisation 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

(b) The organisation has a taxonomy for describing Data Assets, Metadata, and supporting information.

(c) The organisation 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

(a) The organisation must make it easy for Data Users to be able to use and understand information that describes each Data Asset. The organisation must therefore provide Metadata associated with Data Assets and this Metadata must be made available to Data Users independent of the Data Asset. Where practical, this Metadata should be made accessible through a Data Catalogue.

(b) The organisation must treat the Metadata as a Data Asset. When providing Metadata, the organisation must format and structure this Metadata using the latest version, or a subsequent iteration, of a recognised Metadata standard, such as the Dublin Core Metadata Standard.

(c) There is no requirement for the organisation to create Metadata about its Metadata associated with Data Assets.

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

Intended outcome

(e) Where practical, the organisation has a Data Catalogue in a location which is accessible to current and prospective Data Users.

(f) The organisation can demonstrate how it uses the latest version, or a subsequent iteration, of a recognised Metadata standard when utilising and providing Metadata.

(g) The organisation 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 Data Assets by providing supporting information.

Explanation

(a) Throughout the lifecycle of a Data Asset the organisation must make available supporting information that helps prospective and current Data Users to understand the Data Asset. The organisation 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

(b) The organisation can demonstrate that they provide clear and concise supporting information to both prospective and current Data Users, alongside their Data Assets.

(c) The organisation 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 for potential Data Users.

Explanation

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

Intended outcome

(b) The organisation publishes Metadata allowing users to identify the contents of Data Assets.

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

Explanation

(a) The organisation must identify the requirements of Data Users who use, or who wish to use, the Data Assets provided by it as Data Custodian. The organisation must then develop and deliver the project in accordance with the Data Users' requirements.

Intended outcome

(b) The organisation can demonstrate its process for gathering information on user needs from its Data Users.

(c) The organisation can demonstrate its process for assessing what is needed to meet the needs of Data Users, and whether the project activities align with their needs.

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

Explanation

(a) Where the organisation reasonably expects the Data Users' application of Data Assets for which it is the Data Custodian to deliver a net benefit for stakeholders, the organisation 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.

(b) Where data quality issues are identified the organisation must ensure that these issues are logged, considered and rectified as soon as practicable.

Intended outcome

(c) The organisation 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.

(d) The organisation 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

(a) The organisation must enable interoperability, between the Data Assets for which it is Data Custodian and Data Assets of other similar organisations as a minimum standard.

(b) When the organisation 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 organisations. There must also be sufficient information to align to Data Assets from other industries.

(c) The organisation 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 organisation'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

(d) The organisation can demonstrate how they have delivered, and will be delivering, the project in a way that will enable interoperability.

(e) The organisation can demonstrate how they have delivered their projects in a way that Data Users can easily join their Data Assets with Data Assets of other organisations.

9. Protect Data Assets and systems in accordance with current regulations and legislation relating to cyber security.

Explanation

(a) The organisation must ensure adherence to relevant regulations and legislation in relation to cyber security.

Intended outcome

(b) The organisation can demonstrate how its project is delivered in compliance with the current regulations and legislation relating to cyber security.

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

Explanation

(a) When Data Assets are not required by the organisation, the organisation must ask stakeholders whether they consider that the Data Assets could create a future benefit if archived. The organisation must archive Data Assets when, taking account of stakeholders' views it determines that the storage will be a net benefit to stakeholders.

(b) When archiving, the organisation 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.

(c) The organisation 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

(d) The organisation can demonstrate its process for archiving Data Assets and how this process has been built in consultation with Data Users.

(e) The organisation 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

(a) The organisation 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.

(b) The organisation 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.

(c) Where a sensitivity is identified with the Data Assets, their associated Metadata and Software Scripts used to process Data Assets, the organisation 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 organisation 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.

(d) The organisation 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 stakeholders to do so.

(e) The organisation 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.

(f) The organisation 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.

(g) The organisation 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

(h) The organisation can demonstrate that it applies the Open Data Triage process as defined in The Guidance.

(i) The organisation 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.

(j) The organisation 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.

12. Ensure that Data Assets, Metadata and Software Scripts are collected, used or shared with due consideration for Data Ethics.

Explanation

(a) The organisation must ensure that where Data Assets, Metadata or Software Scripts are collected, used or shared by the organisation, ethical implications of their collection, use or sharing are assessed.

(b) The organisation must document the ethical assessment, as well as any changes or mitigations made to the collection, use or sharing of Data Assets, Metadata and Software Scripts, as a result of the assessment.

(c) The organisation's supporting information regarding a Data Asset must include any relevant information for its ethical use or sharing by Data Users.

Intended Outcome

(d) The organisation can demonstrate that Data Assets, Metadata and Software Scripts held by the organisation will be collected, used or shared ethically by the organisation.

(e) The organisation also provides sufficient information to Data Users to ensure they can also use or share the Data Assets, Metadata and Software Scripts ethically.

(f) Stakeholders trust that Data Assets are respected by the organisation and that Data Ethics risks are avoided.

(g) Derived insights from Data Assets should be created and presented in such a way that ethical considerations of the Data Assets are understood by stakeholders.

4. Available Support

The Principles
The principles of The Guidance are based on the Ofgem's Data Best practice guidance with one additional principle on Data Ethics. The best available supporting information is Ofgem's Data Best Practice supporting information document. At the time of writing, that version is hosted on Ofgem's website. Other available support on open data licence, data triage, data sharing and data ethics are listed below.

Open data licences guidance
Where an organisation, consortium, individual or project group is following The Guidance, they will be expected to conduct Open Data Triage and make their Open Data available, where applicable, using open data licences. An open data licence allows Data Users to use, copy, publish and redistribute data and with the appropriate types of licence, it encourages Data Users to acknowledge the source of information.

There are a number of available open data licences to choose from depending on the purpose and nature of work. The Guidance sets an expectation to use either the Creative Commons Attribution Licence or the Open Government Licence. The Creative Commons licence chooser beta is a tool that helps Data Controllers select the appropriate licence for their work under the Creative Commons umbrella. Further information on Open Government Licences are provided for funding recipients use.

Data triage guidance
Where an organisation, consortium, individual or project group are required through contractual arrangements to follow The Guidance they will be expected to conduct Open Data Triage on Data Assets that have been created during the course of funded work. Definitions on Data Assets and Open Data Triage can be found in the glossary.

Ofgem's Data Best Practice Guidance Supporting Information, specifically pages 56 - 78 provide extensive guidance on how to conduct Open Data Triage. The Energy Networks Association also provides guidance in their Data Triage Playbook. The Centre for the Protection of National Infrastructure (CPNI) also has guidance on a Triage Process for the publication or disclosure of information.

In considering the data triage progress, the Data Spectrum is an advocacy tool that helps to understand the language of data and openness. The left hand side of the spectrum considers data that is meant to be accessed internally. This type of data is often sensitive and involves personal information, such as customer profiles, the sharing of which may put individuals at risk. As data moves along the Data Spectrum from left to right an increasing number of people will have access to the data.

Subject to the Open Data Triage, all data gathered or created in the course of a project should be made openly available on a publicly accessible website where

possible. Some examples of publicly accessible websites include the Open Energy Platform, Open Net Zero, UKERC energy data centre and Data Mill North.

Data sharing guidance and model data sharing agreements
Organisations can make use of the Data Ecosystem Mapping tool to map out the data ecosystem. This consists of the data infrastructure, the people, communities, and organisations that benefit from the value created by it. It helps explore the variety of ways in which data creates value, illustrate the different actors in a data ecosystem, and show how value is exchanged across it.

The Information Commissioner's Office (ICO) created a code of practice on data sharing to help organisations comply with data protection law. While the code primarily focuses on personal information, it provides a comprehensive checklist of considerations that organisations should think about before sharing data. The data sharing agreement template attached to the code is also a good model to make reference from when taking the decision to share data. Another guidance would be the Designing data sharing agreements: A checklist, which takes users through steps to consider data sharing needs, whether a data sharing agreement is needed and if so, what to include.

Innovators may also look into successful case studies of data sharing in other industries. Open Banking, for example, is the process of banks and other financial institutions releasing certain data like product descriptions and branch locations in a standardised format, as well as the secured sharing of more sensitive consumer data to trusted third party organisations. The Operational Guidelines, which supports the design of an effective and compliant open banking interface, comes with a checklist which contains a list of questions that help service providers to assess their conformance against key criteria identified in the guideline.

Data ethics assessment
It is fundamental to consider Data Ethics when collecting, accessing, using and sharing Data Assets. It is a particularly important aspect to deliver for the public when project funding comes from public funding. For this reason a new principle on data ethics, Principle 12, has been added to The Guidance.

There are various tools available to help innovators understand and assess Data Ethics at governmental, organisational and project levels. The Central Digital and Data Office (CDDO) promulgated the Data Ethics Framework to guide appropriate and responsible data use in government and the wider public sector. The framework, split into overarching principles and specific actions, helps public servants and arms-length bodies to understand Data Ethics considerations and address any risks concerned with their projects.

At the organisational level, the Data Ethics Maturity Model is a tool to help assess how widely embedded data ethics culture and practices are across an organisation. This is suitable for anyone who collects, uses and shares data. The model provides a skeleton to consider current performance and future ambitions in the team with regard to data ethics, and allows development of action plans to improve existing

shortcomings. It features six different aspects of data culture including organisational governance and internal oversight, skills and knowledge, data management risk processes, funding and procurement, stakeholder and staff engagement, as well as legal standing and compliance.

The Data Ethics Canvas and Consequence Scanning are complementary tools that help illuminate and manage ethical issues and mitigation responses at any stage in a data project by asking questions about the use of data in the project. They provide a framework to develop ethical guidance that suits any context, regardless of the size or scope of the project.