

DATA BEST PRACTICE CONSULTATION RESPONSE

GREG JOHNSTON

SENIOR DIGITAL AND DATA CONSULTANT

GREG.JOHNSTON@ES.CATAPULT.ORG.UK

APRIL 2023



No IP restrictions apply.

CONTENTS

1.	OVERVIEW	1
2.	QUESTIONS AND ANSWERS	3
2.1.	Q1. Do you agree with our proposal to implement a structural change to DBP Guidance, introducing intended outcomes for each principle? If not, how do you suggest we could clarify the aim of each principle?	3
2.2.	Q2. What are your views on the proposed wording of our intended outcomes for each principle in DBP Guidance?	3
2.3.	Q3. What are your views on our proposal to require the use of Dublin Core as the Metadata standard for companies obligated under DBP Guidance?	3
2.4.	Q4. If you do not agree with this proposal, are there alternative Metadata standards that should be utilised by licensees instead?	3
2.5.	Q5. If you are a licensee required to comply with DBP Guidance, can you provide a timescale for the implementation of the proposal to adopt Dublin Core as your Metadata standard?	3
2.6.	Q6. What are your views on our proposal to require the use of the Creative Commons Attribution Licence or the Open Government Licence as the standard open data licence for companies obligated under DBP Guidance?	4
2.7.	Q7. If you do not agree with this proposal, can you suggest alternative open data licences to be utilised as a common open data licence?	5
2.8.	Q8. If you are a licensee required to comply with DBP Guidance, can you provide a timescale for the implementation of the proposal to adopt the Creative Commons Attribution Licence or the Open Government Licence as your open data licence?	5
2.9.	Q9. What are your views on our proposal to require licensees to create and publish a Data Catalogue of their Data Assets?	5
2.10.	Q10. Do you agree with our proposed position on treating aggregated smart meter consumption data as Energy System Data?	5
2.11.	Q11. What are your views on our position that this Data Asset should be published in a non-interoperable fashion by 14 October 2023, if the appropriate security controls are in place?	5
2.12.	Q12. What are your views on our proposal that DNOs collectively determine an interoperable methodology by 28 February 2024, for publishing aggregated smart meter consumption data?	6
2.13.	Q13. What are your views on our proposal that licensees treat Data Assets associated with flexibility market operation as Presumed Open?	6
2.14.	Q14. Do you foresee any specific barriers to treating Data Assets associated with flexibility market operation as Open Data?	6

DISCLAIMER

This document has been prepared by Energy Systems Catapult Limited.

All information is given in good faith based upon the latest information available to Energy Systems Catapult Limited. No warranty or representation is given concerning such information, which must not be taken as establishing any contractual or other commitment binding upon the Energy Systems Catapult Limited or any of its subsidiary or associated companies.

1. OVERVIEW

Energy Systems Catapult (ESC) welcomes the opportunity to respond to the Data Best Practice Guidance consultation.

ESC was set up to accelerate the transformation of the UK's energy system and ensure UK businesses and consumers capture the opportunities of clean growth. ESC is an independent, not-for-profit centre of excellence that bridges the gap between industry, Government, academia, and research. We take a whole systems view of the energy sector, including in policy design and implementation, helping us to identify and address innovation priorities and market barriers, to decarbonise the energy system at the lowest cost.

ESC is supportive of the changes suggested by Ofgem, with additional comments to be found in direct response to each question set out in this document.

Additionally, we would like to make the case for an additional principle around Data Ethics. We have worked with the Open Data Institute to develop this new principle and have socialised that concept to the industry with posts that can be found on the websites of [the ODI](#), [Energy Systems Catapult](#), & [Utility Week](#) as well as a number of linkedin and twitter posts by both the ODI and ESC. The blog also appeared in the Utility week newsletter on Thursday April 13th.

We believe that Data Ethics is an important consideration for data practitioners and including a principle will play a role in supporting a just transition of the energy sector. Our suggested new principle, written in the style of Data Best Practice guidance, is set out on the following page. While not set out in the scope of the consultation itself, we believe if other respondents, particularly energy networks respond in support of this new principle, it should be included. Inclusion within the supporting information may be a palatable first step now, before consulting on formally including it within the guidance at a later stage, though we do hope to see it taken up at this opportunity.

Please note, the draft for the Data Ethics principle found on the following page is slightly different to that which was posted in the links that were socialised on/around the 12th of April. This is due to ESC working with a researcher to try and iterate and improve upon the principle, which occurred after agreeing the text with the ODI as part of the Innovate UK policy work, and during the process of agreeing language for the blog. For those reasons, our recommendation within this consultation response is slightly different.

[Definition] 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.

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

Explanation

- a. The licensee must ensure that where Data Assets, Metadata or Software Scripts are collected, used or shared by the licensee, ethical implications of their collection, use or sharing are assessed.
- b. The licensee 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 licensee'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 licensee can demonstrate that Data Assets, Metadata and Software Scripts held by the organisation will be collected, used or shared ethically by the licensee.
- e. The licensee provides sufficient information to Data Users to ensure they can use or share the Data Assets, Metadata and Software Scripts with due consideration of ethics.
- f. Stakeholders trust that the licensee is collecting, using or sharing data in an ethical way.
- 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.
- h. Licensees use or develop ethical frameworks around the collection, use and sharing of Data Assets, Metadata and Software Scripts.

2. QUESTIONS AND ANSWERS

2.1. Q1. DO YOU AGREE WITH OUR PROPOSAL TO IMPLEMENT A STRUCTURAL CHANGE TO DBP GUIDANCE, INTRODUCING INTENDED OUTCOMES FOR EACH PRINCIPLE? IF NOT, HOW DO YOU SUGGEST WE COULD CLARIFY THE AIM OF EACH PRINCIPLE?

We believe the change is helpful and the suggested language under each principle should provide sufficient clarification for those following the guidance.

2.2. Q2. WHAT ARE YOUR VIEWS ON THE PROPOSED WORDING OF OUR INTENDED OUTCOMES FOR EACH PRINCIPLE IN DBP GUIDANCE?

Principle 5's current wording of the intended outcome suggests that each network must own and operate a data catalogue. Specifically, use of the word "Has" in 3.18 & "their" in 3.19. We suggest amending to allow for network operators to be given the option to create or use a catalogue owned or operated by another industry party, jointly or through some other mechanism.

Under principle 10, we suggest a change in 3.28 from "regular" to "appropriate" or simply deleting the word regular. Otherwise, the use of regular may need further clarification. Given the Digitalisation Strategy and Action Plan guidance, there is sufficient information for networks to take steps for stakeholder engagement, meaning we believe 3.28 doesn't need any level of specificity.

2.3. Q3. WHAT ARE YOUR VIEWS ON OUR PROPOSAL TO REQUIRE THE USE OF DUBLIN CORE AS THE METADATA STANDARD FOR COMPANIES OBLIGATED UNDER DBP GUIDANCE?

ESC is very supportive of this step to require Dublin core as the metadata standard. Further thinking may be required to manage updates and deviations by industry bodies with some sort of working group around the standard's use in the energy context. A broader policy question around the governance of data standards in the sector is emerging, and could do with an industry wide view on how to resolve.

2.4. Q4. IF YOU DO NOT AGREE WITH THIS PROPOSAL, ARE THERE ALTERNATIVE METADATA STANDARDS THAT SHOULD BE UTILISED BY LICENSEES INSTEAD?

Not applicable.

2.5. Q5. IF YOU ARE A LICENSEE REQUIRED TO COMPLY WITH DBP GUIDANCE, CAN YOU PROVIDE A TIMESCALE FOR THE IMPLEMENTATION OF THE PROPOSAL TO ADOPT DUBLIN CORE AS YOUR METADATA STANDARD?

Not applicable.

2.6. Q6. WHAT ARE YOUR VIEWS ON OUR PROPOSAL TO REQUIRE THE USE OF THE CREATIVE COMMONS ATTRIBUTION LICENCE OR THE OPEN GOVERNMENT LICENCE AS THE STANDARD OPEN DATA LICENCE FOR COMPANIES OBLIGATED UNDER DBP GUIDANCE?

We think this is a sensible proposal but do caution that the interoperability between licences is full. While the legal interoperability between licenses is noted by OGL on their website as being interoperable with Creative Commons¹, it is not clear if this is entirely accurate. A recent blog by Yui-Shing Pang at UKPN² noted a challenge to this thinking – summed in the quote on the right.

We are of *the opinion it would be useful to drop the option of using OGL altogether*, as CC-BY 4.0 is the more commonly used standard and therefore creating a bigger “silo” in the first instance. Our understanding is that CC-BY 4.0 license use would comply with the licence attribution requirements of OGLv3.0, but CC-BY 4.0 attribution requirements may not be satisfied if an OGLv3.0 were used. This means that were an innovator to use data with sources from both licence regimes, they would typically be required to follow through with using CC-BY 4.0 licensing for their purposes. If proprietary licence terms were placed upon OGLv3.0 licences by a network, or any other innovator in a data chain, this could create challenges for attribution and use.

Our opinion that CC-BY 4.0 would be a preferable standard is supported by the notion that OGL licences are designed by use for public sector organisations, and beyond a possible reluctance for licensees to use OGLv3.0, the OGLv3.0 licence also enables proprietary licence terms on onward distributions of the data, i.e., an organisation can set restrictions on forward use in a way that may not align with the spirit or intent of Data Best Practice. This could be done by, for example, imposing proprietary licences on the data through OGLv3.0 which could give the appearance of ownership rights over said data. Finally, CC-BY 4.0 licence terms deal specifically with database issues, where these are only dealt in general terms under OGLv3.0.

We also believe that Ofgem should ensure that the class of non-commercial creative commons licences are not used, such as CC-BY-NC 4.0. Doing so would go against the presumed open principle and contradict the Open Data definition in the glossary. While it may be sufficiently understood by practitioners that NC licences aren’t open data, it is likely that clarity would need to be provided by Ofgem to licensees.

Even less certain is that; where open data in the future may include non-open standards, such as CIM (*depending on the GB implementation and associated licensing*) a situation may arise where open data may inadvertently become a derivative work³ of the non-open standard. A legal review

However, OGL v3.0, in its unchanged form, contains a single clause which presents a point of incompatibility, making the associated dataset less open:

“This licence is governed by the laws of the jurisdiction in which the Information Provider has its principal place of business, unless otherwise specified by the Information Provider.”

¹ <https://www.nationalarchives.gov.uk/information-management/re-using-public-sector-information/uk-government-licensing-framework/open-government-licence/>

² <https://www.linkedin.com/pulse/open-licences-yiu-shing-pang/>

³ https://en.wikipedia.org/wiki/Derivative_work

by Ofgem, DESNZ or another party may be appropriate in the medium term to provide clarity on if this is a tangible risk, and what steps can be taken to mitigate it.

In summary, we believe CC-BY-4.0 licences should be the standard taken forward by Data Best Practice guidance and that further work should be done to understand the risks associated with derivative works being inadvertently created.

2.7. Q7. IF YOU DO NOT AGREE WITH THIS PROPOSAL, CAN YOU SUGGEST ALTERNATIVE OPEN DATA LICENCES TO BE UTILISED AS A COMMON OPEN DATA LICENCE?

See notes in Q6 answer.

2.8. Q8. IF YOU ARE A LICENSEE REQUIRED TO COMPLY WITH DBP GUIDANCE, CAN YOU PROVIDE A TIMESCALE FOR THE IMPLEMENTATION OF THE PROPOSAL TO ADOPT THE CREATIVE COMMONS ATTRIBUTION LICENCE OR THE OPEN GOVERNMENT LICENCE AS YOUR OPEN DATA LICENCE?

Not Applicable.

2.9. Q9. WHAT ARE YOUR VIEWS ON OUR PROPOSAL TO REQUIRE LICENSEES TO CREATE AND PUBLISH A DATA CATALOGUE OF THEIR DATA ASSETS?

ESC is supportive of this proposal, as it aligns with the work done under the Energy Data Visibility Project (EDVP), Open Energy more broadly and supports the recommendations of the Energy Data Taskforce. Our suggested amendments to the intended outcomes section of the principle five change are noted in our response to question two of this consultation. We believe the purpose of this principle should be to develop a catalogue, or a suite of data catalogues. As it is currently written, we interpret a decision, or an inadvertent outcome is found, that each licensee will have their own individual catalogue, with inference that this would be hosted locally in each instance – which may not be the optimal solution. The work undertaken by ESC, Arup and the University of Bath in the feasibility of a digital spine, may also offer an opportunity for different approaches to data catalogues being undertaken.

2.10. Q10. DO YOU AGREE WITH OUR PROPOSED POSITION ON TREATING AGGREGATED SMART METER CONSUMPTION DATA AS ENERGY SYSTEM DATA?

Yes, this is a very positive step. The availability of aggregated smart meter data will enable a number of use cases from commercial propositions to academic research and should be made available at the earliest opportunity.

2.11. Q11. WHAT ARE YOUR VIEWS ON OUR POSITION THAT THIS DATA ASSET SHOULD BE PUBLISHED IN A NON-INTEROPERABLE FASHION BY 14 OCTOBER 2023, IF THE APPROPRIATE SECURITY CONTROLS ARE IN PLACE?

We are very supportive of this timetable and hope to see the networks prioritise this work – along with partners in Ofgem in supporting any necessary changes in the Smart Meter Data Privacy Plans.

We note however that there are risks of identification of households where DNO's publish data ahead of a standardised approach being developed. Where a deadline is set (October 14) and networks are satisfied their security and privacy controls are in place we would expect them to publish data. If this publication is *simply* the aggregation of data as set out in the current data privacy plans, then there is no mathematical certainty that the data is secure. A residual risk, not

currently precisely articulated by academics we speak to, remains. The lack of articulation of that risk poses a challenge, as we are not confident of the size of that risk to an individual's privacy.

Additionally, there are theoretical risks of bad faith actors using aggregated data, and other data sources such as census data to de-aggregate and identify an individual household signal. Car registration data, in conjunction with aggregated smart meter data which identifies feeders, may be able to identify an EV owner within the signal, for example. We are of the impression that the required technical skill barriers, along with no clear incentive to do so, may prevent bad faith actors from doing this. However, the development of large language models (LLM's) and the associated technical capabilities it provides may make the disaggregation an easier challenge to overcome if someone was sufficiently motivated and incentivised to identify an individual household.

We believe that a **differential privacy** approach, as explored in a paper⁴ from the energy futures lab at Imperial, would be an appropriate security mechanism for smart meter data and should form the basis of 'an interoperable methodology' as described in Q12. More information on differential privacy as an approach can be found by looking at the U.S. Census Bureau's brief⁵ on its use for U.S census data. We also note another challenge that sits outside of the scope of DBP guidance. Where individuals use home energy management systems (HEMS) and have commercial arrangements with providers, the profile of their usage will be known. Depending on specific contractual arrangements, this means that that profile data can be bought or sold to different providers.

Depending on these contracts, and the proliferation of HEMS over time, privacy may be reduced. For example, under one feeder with a DNO aggregation of 'n' customers, if 'n-1' have 'sold' their profile data via arrangements with HEM providers then everyone can be immediately known even if not all have 'sold' their data. The interaction with HEMS will need to be explored further, though it may be out of scope for the regulator's responsibilities.

2.12. Q12. WHAT ARE YOUR VIEWS ON OUR PROPOSAL THAT DNOS COLLECTIVELY DETERMINE AN INTEROPERABLE METHODOLOGY BY 28 FEBRUARY 2024, FOR PUBLISHING AGGREGATED SMART METER CONSUMPTION DATA?

Please see our commentary in Q11 around differential privacy.

2.13. Q13. WHAT ARE YOUR VIEWS ON OUR PROPOSAL THAT LICENSEES TREAT DATA ASSETS ASSOCIATED WITH FLEXIBILITY MARKET OPERATION AS PRESUMED OPEN?

Yes, we believe that this is a good approach. In the long term, it may be sensible to include a licence condition within the scope of the SSES consultation work by DESNZ to compel those proposed to be licensed to follow data best practice. A similar case should also be made for future licence conditions relating to Ofgem's regulation of heat networks given the electricity loads that may be used to operate these networks, where heat pump technologies are utilised, for example.

2.14. Q14. DO YOU FORESEE ANY SPECIFIC BARRIERS TO TREATING DATA ASSETS ASSOCIATED WITH FLEXIBILITY MARKET OPERATION AS OPEN DATA?

No, this is a sensible move and should be undertaken bearing in mind the licensing approach discussed in our answer to question six.

⁴ <https://www.imperial.ac.uk/energy-futures-lab/reports/briefing-papers/paper-9/>

⁵ <https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-03.pdf>

OUR MISSION

**TO UNLEASH INNOVATION
AND OPEN NEW MARKETS
TO CAPTURE THE CLEAN
GROWTH OPPORTUNITY.**

**ENERGY SYSTEMS CATAPULT
7TH FLOOR, CANNON HOUSE,
18 PRIORY QUEENSWAY,
BIRMINGHAM, B4 6BS.**

**ES.CATAPULT.ORG.UK
@ENERGYSYSCAT**