Consultation Response: Securing Open Data in Energy

Respondent: Daramola Joseph Omoyele
Capacity: Independent Data Analyst and Research Analyst
Location: Colchester, United Kingdom
Website/Portfolio: daramolajo.co.uk
Response status: Non-confidential

1. Introduction

I welcome the opportunity to respond to this consultation. I am responding in a personal professional capacity as a Data Analyst and Research Analyst with experience in statistical modelling, data quality assurance, dashboard development, GDPR-aware data handling, and public policy research. My work has involved analysing survey, administrative, operational, and financial datasets, building decision-support tools, improving data accuracy, and communicating technical findings to non-technical stakeholders.

My response focuses on the practical conditions required for open energy data to remain useful, secure, machine-readable, auditable and accessible to smaller innovators, researchers, civic technology users, local authorities, and policy analysts.

My central position is that Ofgem should adopt the Hybrid Model but strengthen it with mandatory schema validation, published metadata standards, compulsory aggregation risk testing, auditable triage records, defined data quality service levels, and explicit safeguards to prevent unnecessary barriers for smaller data users. The aim should not be to publish less data. The aim should be to publish the right data, at the right level of granularity, with the right metadata, through a controlled process that protects critical infrastructure while preserving the innovation value of open data.

2. Question 1: How open data supports products, services, or efficiency savings

I have not built a commercial energy product directly from DBP datasets. My contribution is based on practical data analysis, data governance, dashboard development, and evidence-based policy reporting.

What I can offer is a professional perspective on the conditions under which openly published data creates measurable value. In my own work, open and administrative datasets have supported dashboard development for faster management decisions, trend analysis and early identification of operational risks, evidence-based reporting for senior stakeholders, improved data reconciliation and reduced manual reporting overhead, and regression modelling and scenario analysis using R and Python.

Applying that experience to the energy context, I would draw particular attention to four use cases where open data is most likely to create durable public value.

The first is local energy planning. Local authorities and analysts need reliable data on consumption patterns, network constraints, connection capacity, and local demand trends. Without accessible datasets, local area energy planning becomes slower, more expensive, and more dependent on private access arrangements that smaller or less well-resourced authorities cannot easily navigate.

The second is fuel poverty targeting. Aggregated and privacy-preserving energy data can help identify areas where households may face affordability pressure or poor insulation outcomes. I am conscious

that this use case sits close to a GDPR boundary, and I address that directly later in this response. The value here is not in exposing individual household data. The value is in enabling area-level targeting of support, grants and retrofit planning where it is most needed.

The third is network investment prioritisation. Open network data can help analysts identify where constraints, demand growth and low-carbon technology adoption may require earlier intervention, which can support better prioritisation of capital spending and reduce poorly targeted infrastructure build.

The fourth is innovation by smaller organisations. Smaller firms, researchers and civic technology developers do not usually have the market power or resources to negotiate bespoke data access arrangements. Open data gives them a fair route to test ideas, build prototypes and challenge larger incumbents. This is the use case most at risk if the chosen model creates administrative barriers that larger organisations can absorb but smaller ones cannot.

For these benefits to materialise in practice, publication alone is not sufficient. The data must be machine-readable, version-controlled, accompanied by clear metadata, available through stable endpoints, published with data dictionaries, updated on a predictable schedule, licensed clearly and validated before release. Poor data quality creates hidden costs. Analysts spend significant time cleaning column names, reconciling inconsistent formats, interpreting missing fields and checking whether apparent changes reflect real-world shifts or inconsistent publication practices. Good data governance reduces those costs, and Ofgem should treat data quality as a substantive policy objective, not an operational afterthought.

I would also encourage Ofgem to commission a structured evidence harvest from DBP-obligated licensees before the statutory consultation, requesting quantified benefit estimates rather than illustrative examples. A more robust counterfactual would strengthen the analytical foundation for any reduction in the scope or granularity of open publication.

3. The balance between risk and growth

I agree that Ofgem should not reverse the principle of presumed open data. A retreat from open publication would reduce market transparency, increase information asymmetry, weaken independent analysis, and make it harder for smaller organisations to participate in energy innovation. The consultation document itself acknowledges at paragraph 2.2 that the potential harm of withdrawal outweighs the possible risk, and I agree with that assessment.

However, the risk environment has changed in ways that require a more structured approach to publication decisions. I note in particular paragraphs 2.12 to 2.14 of the consultation, which set out the aggregation risk problem clearly. The increased use of artificial intelligence, machine learning and geospatial analytics means that datasets that appear low-risk in isolation may become sensitive when combined with other open sources. This is a practical risk that any credible triage process should address systematically.

In my view, the issue is not whether data should be open or secure. The issue is how Ofgem can keep data open while applying stronger controls where the risk is higher. Ofgem should retain the presumed open principle but require a structured test before publication is confirmed. In practical terms, that test should ask: what harm could arise if this dataset is combined with other available data; does the dataset expose infrastructure location, operational dependency, timing patterns, or network vulnerability; can the same public benefit be achieved through aggregation, temporal delay, geographic smoothing, or field suppression; what level of granularity is actually necessary for

legitimate users; and has the risk decision been documented in a way that can be audited. These questions should be embedded in the triage guidance, not left to individual judgement.

4. The scoring criteria

I broadly agree with the five criteria used in the options analysis: ownership and accountability, data security, data quality, cost, and complexity of change.

I have two observations on methodology. First, the current approach treats all five criteria as equally weighted when scores are summed. This is a reasonable starting point, but it is methodologically weak for a decision of this significance. Data security and data quality are not equivalent in policy consequence to the complexity of change. I would recommend that the statutory consultation either apply explicit weights to the criteria or include a sensitivity analysis showing how the preferred option would rank under different weighting assumptions. This would make the final decision more defensible, particularly where the scores between models are close.

Second, the scoring does not distinguish between short-run and long-run performance. The Educational Model performs well on cost and complexity now but may generate higher costs over time through inconsistent triage outcomes, regulatory enforcement action, or a security incident. A lifecycle perspective would improve the comparability of the three options.

I would also recommend adding a sixth criterion: access equity and innovation usability. This would assess whether the chosen model allows small businesses, researchers, local authorities, charities, civic technology users, and independent analysts to access open data without excessive administrative or financial barriers. A model could score well on security and governance but still weaken innovation if it introduces slow approval processes, unclear registration rules, poor API documentation, inconsistent metadata, or restrictions that in practice favour larger incumbent organisations. Measuring this dimension explicitly would help Ofgem ensure that securing open data does not inadvertently close it.

In support of that criterion, I recommend that Ofgem track the following operational indicators after implementation: average time from triage submission to publication; percentage of datasets passing schema validation at first submission; percentage of published datasets with complete metadata; number of datasets reclassified from open to shared or closed following aggregation risk review; number of user-reported data quality issues and average resolution time; API uptime; availability of bulk download; and number of active users disaggregated by organisation type, including whether smaller or non-incumbent users are being reached.

5. The Centralised Model

The Centralised Model has a genuine security advantage. A single function applying consistent standards is better placed to identify aggregation risks across multiple datasets, and a named risk owner with clear governance accountability is preferable to distributed responsibility across licensees with variable capacity and risk appetite.

My concern is operational. A fully centralised model could become a bottleneck if the Digitalisation Coordination Function does not have sufficient technical depth, subject-matter expertise and escalation capacity from the outset. There is also a governance risk that licensees begin to treat triage as something done externally rather than a core responsibility of the organisation that understands its own data best. That diffusion of ownership could introduce a different category of risk.

My view is that the Centralised Model should not be the first implementation model. It may be appropriate for the most sensitive categories of Energy System Data once the Digitalisation Coordination Function has demonstrated governance maturity and service-level performance. For now, the operational dependencies are too significant and the timeline too uncertain to make it the primary vehicle for reform.

I agree with Ofgem's scoring on ownership (3), data security (5), and data quality (4). I would note that the cost score of 1 and complexity score of 1 may improve materially once the Digitalisation Coordination Function is designated and a proper business case is developed. Centralised infrastructure frequently delivers economies of scale that are not visible at the design stage.

6. The Hybrid Model

I support the Hybrid Model. It combines licensee knowledge of individual datasets with central validation, improved consistency, a single publication route, and better visibility of aggregation risk across the system. It is also better aligned with mature data governance practice, where data owners remain accountable for what they produce while central functions set standards, validate quality and monitor risks.

I set out below the minimum design features I consider necessary for the model to function as intended.

Accountability must be clearly split. The licensee should remain accountable for data accuracy, the initial sensitivity assessment, documentation of assumptions, identification of known risks, application of mitigation measures, and correction of data quality issues. The Digitalisation Coordination Function should be accountable for validation rules, schema enforcement, aggregation risk checks, publication controls, platform reliability, metadata quality standards, user access monitoring, and escalation where validation fails. This split prevents a situation where no organisation clearly owns the risk, which is the failure mode most likely to occur when responsibility sits between two parties.

Schema validation should be mandatory before publication. Datasets should pass automated checks covering required fields, accepted data types, geographic resolution, timestamp format, missing values, duplicate records, unit consistency, column naming conventions, metadata completeness, licence information and stated update frequency. Datasets that fail validation should not be published until the issue is corrected or formally justified. This is standard practice in mature data pipelines, and regulated Energy System Data should be subject to the same level of discipline.

Metadata should be treated as core infrastructure, not administration. Each published dataset should include the dataset owner, publication date, update frequency, licence, geographic coverage, time period covered, a data dictionary, methodology notes, known limitations, sensitivity classification, mitigation applied, version history, and a contact route for data quality queries. Without this information, data users cannot assess reliability, comparability, or appropriate use. Metadata gaps also create analytical errors that are difficult to detect, particularly when datasets are combined.

Aggregation risk testing should be compulsory before open classification. This means checking whether a dataset becomes sensitive when combined with other published energy data, geospatial infrastructure data, property-level data, demographic data, weather data, or commercial location datasets. The risk assessment should consider whether the combined output could reveal critical infrastructure vulnerability, single points of failure, operational dependencies, exploitable timing

patterns or site-level sensitivity. Where risks are identified, mitigation should be required before publication. That mitigation could take the form of aggregation, temporal delay, geographic smoothing, suppression of high-risk fields, banding, redaction or differential privacy techniques, depending on the nature of the risk.

I note that the consultation at paragraphs 2.12 to 2.14 acknowledges that aggregation risk is covered in existing guidance but still presents complexities without easy solutions. I agree that there is no simple solution. But making aggregation risk testing a compulsory and documented step in the Hybrid Model's validation process, rather than a consideration to be weighed, would at minimum create a consistent and auditable record of how these decisions are being made.

Every triage decision should have an audit trail. The record should show what was assessed, who assessed it, when it was assessed, what risks were identified, what mitigation was considered, why the chosen publication route was selected, whether the dataset was classified as open, shared, or closed, and whether central validation changed the licensee's initial assessment. This does not mean publishing sensitive details. It means Ofgem should require a non-sensitive transparency record so that stakeholders can see that triage decisions are systematic and consistent over time.

Rejected datasets should follow a defined workflow. Where the Digitalisation Coordination Function rejects a dataset, the licensee should receive a clear statement of the reason, an agreed correction deadline, and a route to escalation where the decision is disputed. The consultation at paragraph 3.27 identifies that governance processes will be needed to adjudicate situations where a licensee triages data as publishable but validation fails and notes this without resolving it. I recommend that the statutory consultation include a specific proposal for a transparent, time-limited dispute resolution mechanism with defined criteria and an audit trail. Without this, publication delays could become indefinite and the process could be perceived as arbitrary.

7. The Educational Model

The Educational Model is not sufficient as the primary policy response. Training and practitioner courses are useful and should be retained as a supporting layer within any model. But training alone will not produce consistent data quality, uniform metadata standards, or systematic aggregation risk assessment across multiple publishers with different levels of digital maturity. It also does not create a single validation route or a central quality check, which means variation in practice between licensees is likely to persist.

The model's scored performance of 13 reflects this accurately. My recommendation is that the Educational Model's outputs, updated triage guidance, standardised documentation and practitioner training be incorporated into the Hybrid Model as baseline requirements rather than adopted as a standalone approach.

8. Minded-to position and concluding recommendations

I support Ofgem's minded-to position to adopt the Hybrid Model.
Before finalising the decision, I recommend the following implementation safeguards.

The first is to add access equity and innovation usability as a formal sixth criterion in the options analysis, with the operational indicators set out in section 4 of this response used to measure performance after implementation.

The second is to require mandatory schema validation as a condition of publication, with a defined workflow for datasets that fail.

The third is to require standard metadata fields for every published dataset, treated as a publication prerequisite rather than a best endeavour.

The fourth is to make aggregation risk testing a compulsory and documented step before any dataset is classified as open, with clear guidance on mitigation options where risks are identified.

The fifth is to publish annual performance data on validation outcomes, publication timelines, metadata completeness, user access patterns, and datasets reclassified following review. This would allow Ofgem, licensees and data users to assess whether the model is delivering both security and usability over time.

On the cost implications, the consultation notes that network companies under RIIO3 have submitted business cases averaging £15.5 million to meet DBP obligations. The additional requirements I recommend, particularly schema validation, metadata standards, and aggregation testing, carry a cost. My view is that some of these costs are already embedded in existing DBP compliance activities and could be redirected rather than added. Where genuinely new costs arise, the Digitalisation reopener provides a mechanism to address them. The proportionality argument is straightforward: the cost of a security incident or a significant data quality failure affecting the energy system would substantially exceed the investment required to operate a robust validation and governance process.

On the question of access for smaller users, I want to be direct. Registration for platform monitoring purposes is reasonable. But registration should not become a de facto barrier that excludes researchers, local authority analysts, civic technology developers, or independent analysts who lack compliance teams or industry relationships. The open data platform should provide free access to open datasets, clear API documentation, bulk download options, CSV and JSON as minimum formats, clear licence terms, changelogs, archived versions, and transparent publication schedules. These are basic platform features that would help ensure open data remains usable by both large organisations and smaller public-interest users.

Finally, I recommend that the consultation separately consider a risk-tiered approach to publication as a complement to the Hybrid Model's triage process. Low-risk datasets should be published openly with full metadata and API access. Medium-risk datasets should be published at an aggregated level with delayed timing or field suppression where necessary, with stronger metadata and risk notes. High-risk datasets should be classified as shared or closed, with non-sensitive metadata published where safe and access provided through the Data Sharing Infrastructure to accredited users. This tiering would allow the Hybrid Model to maintain the presumed open principle while applying proportionate controls at each level of risk.

Daramola Joseph Omoyele
Independent Data Analyst and Research Analyst
Colchester, United Kingdom

Non-confidential response. I am content for this response to be published in full.