The marginal cost restriction on public sector data re-use

Open Data Manchester workshop — non-attributable readout

Date: 3 September 2026 **Format:** One-hour online session **Participants:** Fifteen, comprising policy officials from the Department for Digital, Culture, Media and Sport who lead on the call for evidence, a policy lead on data sharing from another government department, local government from Greater Manchester and Lancashire, two journalists, academic researchers in data science and research integrity, a civil society open data standards body, and Open Data Manchester.

Open Data Manchester convened this session to help people prepare responses to the government's call for evidence, *The marginal cost restriction on public sector data re-use*, which closes at 11:59pm on 8 September 2026. Officials from the department presented and took questions, and the bulk of the session was open discussion.

This readout is non-attributable. Points are recorded as raised, without attribution to individuals. Statements made by officials are identified as such, because the distinction matters to the record, but not attributed to named people.

Background: the call for evidence

This section is context on the call for evidence. It is not a record of the presentation.

Under the Re-use of Public Sector Information Regulations 2015, public bodies within scope may charge no more than the marginal cost of reproducing, providing and disseminating data they make available for re-use. For digital data that cost is effectively zero, so in practice the regulations require most public sector data in scope to be made available free, as open data under the Open Government Licence.

There are existing exemptions: bodies required to generate revenue covering a substantial part of their public task costs, such as trading funds; bodies required to generate revenue covering collection or dissemination costs; and libraries, museums and archives. Data produced outside a body's public task also falls outside the regulations altogether.

The call for evidence asks whether public bodies should be able to charge above marginal cost for a wider range of data assets, and if so, on what criteria and decided by whom. It follows the Industrial Strategy commitment to develop a Public Sector Data Valuation

Framework, which was published within HM Treasury's Balance Sheet Framework in April 2026, and which sets out that non-open approaches may in some cases be more effective at realising the value of a data asset.

It contains nine questions, grouped around the impact of current rules, changing current rules, design principles, and reform approach.

What officials told the session

Recorded because several points went beyond what the published document says.

This is a call for evidence, not a consultation. Officials were explicit that this is an earlier stage of policy development, that there is no settled policy position in this area, and that a decision not to reform remains a possible outcome.

The objective is cost recovery, not profit. Officials stated that data held by government is already paid for by the taxpayer, and that charging in a way that generated profit would be treated by the Treasury as a secondary tax and is not available to them. The aim described was allowing public bodies not to be out of pocket for making data available.

The barriers as public bodies report them. Two were described. First, incentives: making data available for re-use is generally not part of a body's public task, so bodies absorb the cost and carry the risk — including reputational risk where the data would reveal quality problems — while receiving no benefit. Second, technology: legacy systems, inconsistent standards and interoperability work all cost money.

Scope. Reform would cover only accessible, non-sensitive, non-restricted data. Health and social care data is out of scope, and officials noted that powers to charge for health data already exist and are used. Access rights under freedom of information legislation would be unaffected; the distinction drawn was between the right to *see* data and the right to *re-use and process* it.

Competition law is a live constraint. Officials described government as holding a monopoly over this data, with the consequence that licence terms offered to one organisation must be available to others on the same basis. Freemium tiers or usage-based pricing were mentioned as possible routes to differentiate between a large technology company and a journalist or charity, but officials described themselves as somewhat hamstrung on this point.

No comparative assessment has been done. Asked directly whether any assessment existed of the value of the current system as against a future charging system, officials said no, that this call for evidence is a first step towards it, and that the valuation framework had indicated the area was worth exploring in more detail.

What officials most want in responses. Asked whether responses should address principles or gather evidence of impact, officials were clear that real-world evidence is the priority — particularly evidence of the benefits of the current system and how those benefits might be diminished. It was noted that the published questions may read as principle-focused, and that a response consisting only of concrete examples would be entirely acceptable. Officials also stated that every response has to be read and taken into account.

Two things officials said were still open. Whether it is government's role to add value to data before releasing it, or whether government should release data as it is and let others do that work. And whether the exercise makes sense for smaller data assets at all — the example given was that if standing up a charging mechanism cost in the order of £10,000 to recover small sums per transaction, it would not be worthwhile.

Discussion: points raised

1. Research access and the cost of friction

The strongest sustained theme. Research in data science needs real data, and projects gravitate towards data that is accessible and whose licence permits sharing between collaborators. The UK has generated a considerable body of insight about itself because its public data has been easy to obtain and easy to pass on.

A concrete contrast was offered. Work using open public data on health and vaccine attitudes was possible precisely because the data was available. Work on public procurement had to go through a commercial provider, and the friction involved meant researchers tend not to pursue that kind of application at all. The barrier does not have to be absolute to be decisive; it only has to make one avenue harder than another.

The related point was that open access also delivers transparency and accountability, because anyone can obtain the data and check what has been said about it.

2. Reference data is different in kind

A distinction the consultation does not draw, and one participants thought it needs to: between **reference data**, which identifies the object that other data attaches to, and the substantive data attached to it.

If a body charges for reference data, it is effectively permitting a levy on everything attached to those references. Addresses were given as the case that has been contested for decades, and unique property reference numbers as its current form — fundamental reference data, and consistently among the first things anyone tries to commercialise.

Officials responded that this had been considered but not assessed, that any criteria would need to examine downstream impacts, and that if charging for something like UPRN data produced knock-on effects on other data using those references, they would likely conclude the benefit was not there. They asked for exactly this kind of point to be submitted in writing.

3. The measurement problem, and the asymmetry it creates

Raised as a methodological question. The Balance Sheet Framework acknowledges the difficulty of measuring the wider impacts of open data, while setting out three established methodologies for valuing the data asset itself. The effect is that one side of the comparison has method and the other has acknowledgement.

The 2017 Transport for London study, which put the value of TfL's open data at up to £130m a year, was noted as the last substantial piece of work in this space in the UK.

Officials accepted that valuing data is genuinely contested — that data might be said to have intrinsic value based on quality and provenance, but that in reality value is realised only through use, and that broadening the pool of organisations able to use data magnifies the potential value. They also acknowledged that you never know in advance how data will be used, and that the benefits which followed from TfL opening its data were not foreseen at the time.

4. The cost of running a charging regime

Two questions that the call for evidence does not address.

What would it cost to move from a default where everything is published to one where each asset is weighed, designated, and then managed — access committees, licensing, administration? That bureaucratic cost falls on the public body and would eat into whatever revenue is raised.

And would a charging regime actually deter bad actors? Given several years of large-scale data scraping conducted in disregard of licence terms, a model that says "you may see this data but not re-use it" may constrain the compliant while doing little about the non-compliant.

Officials acknowledged both. On enforcement, they noted the regulator would most likely be the ICO and that resourcing enforcement is a real question. On the underlying trade-off, they suggested the hope would be that charged data is processed and maintained to a higher standard, with data obtained through an access request remaining available but harder to use — which participants noted is itself a description of a two-tier system.

5. Deciding what falls into which bucket

Participants recognised the underlying principle from adjacent debates about high-value datasets — that one might want academic and research access to remain free while charging commercial users. The difficulty identified was not the principle but the sorting: what actually determines which data goes in which bucket, who decides, and on what evidence.

This connects directly to the competition law constraint officials described. If terms offered to one re-user must be available to all, then the sorting cannot be done by who the user is, only by how the data is packaged or how much is used.

6. An alternative model: open unless for profit

Proposed from the floor, by analogy with software licensing that is free for general use and chargeable for commercial use. The default stays open, and payment is triggered by commercial exploitation.

Officials noted it, observed it implies a judgement that for-profit use is in some way less deserving, and mentioned that others have suggested a domestic-versus-international distinction on similar lines. The competition law point above bears on whether either is workable.

7. Datasets that exist in order to be open

Raised in relation to the Bus Open Data Service (BODS), which mandates publication by bus operators and on which practical tools are built and used by local authorities. The observation was that charging for a dataset created expressly to be openly available would defeat the purpose of having created it.

The general form of the point: some public data exists *because* a decision was taken to make it open. Applying a charging framework to that category is different from applying it to data collected for an administrative purpose that happens also to be published.

8. Dependency that is invisible until it breaks

Several participants described consuming government data as an input to something else — a standards body taking government data into its own publication pipeline, local authorities using tools built on national open datasets, newsrooms relying on published government sources precisely because they are trusted and citable.

In each case the dependency is not visible to the publishing body, and the consumer may not learn of a licence change until their pipeline stops working.

Points that may warrant a direct response

The call for evidence closes on 8 September. Officials were explicit that concrete evidence of real-world impact is more valuable to them than commentary on principles, and that a response consisting solely of examples is welcome.

1. Does the framework distinguish reference data — identifiers to which other data attaches — from substantive data? Charging for the former permits a levy on everything attached to it.
2. What is the estimated administrative cost of operating a designation and charging regime, and how does it compare to the revenue anticipated?
3. How will the assessment correct for the measurement asymmetry the Data Valuation Framework itself acknowledges, where the value of open re-use is hard to quantify and the revenue from charging is not?
4. What happens to datasets that exist specifically because a decision was taken to publish them openly, such as those created under statutory open publication duties?
5. How would a charging regime be enforced against non-compliant re-use, given the recent history of large-scale scraping in disregard of licence terms?
6. Given the competition law constraint described, what mechanisms are actually available to ensure affordability for researchers, journalists, charities and small organisations?
7. What evidence exists that revenue raised would be retained and reinvested in data infrastructure, rather than absorbed elsewhere?
8. Has the friction effect on research been assessed — not whether data remains obtainable, but whether researchers will pursue questions that require paid or negotiated access?

Next steps

- Responses are submitted through the online survey: https://www.smartsurvey.co.uk/t/dsitpublicsectordatacfe/
- The call for evidence closes at **11:59pm on 8 September 2026**.
- Officials confirmed there is no requirement to answer every question. A response addressing a single question is acceptable.
- The nine questions are listed in the call for evidence document itself. The survey shows one page at a time, so it is worth drafting answers in a document and pasting them in.
- Open Data Manchester will prepare its own response, informed by this session.
- Participants who did not have the opportunity to raise a point are encouraged to send it on before the deadline.