Falling behind? The UK's high-value public datasets

CFPD, December 2023

This report introduces the EU's recent high-value datasets directive, compares equivalent high-value datasets in the UK, and makes recommendations on how the UK can improve access to data to support high-growth companies. It is the first in a series of notes covering government data sharing, data standards, and other topics. Comments are welcome to contact@centreforpublicdata.org.

Summary

- In 2022, the EU specified certain high-value, non-personal datasets that member states must make freely available, including data on addresses, the environment, and mobility. Other countries have taken similar measures to support growth.
- However, the UK's current national data strategy does not include measures to improve access to high-value public-sector data, and a recent official review warned that access to data remains a barrier for innovation.
- We audited the UK equivalents of the EU high-value datasets, and found that the UK suffers from patchy data availability - data on addresses, buildings, land use and weather remains complex and expensive to access. This risks making the UK less attractive to businesses, and slowing down public service delivery.
- We recommend the government should:
  - Prioritise improving access to high-value datasets as part of its industrial strategy, focussing on non-personal data;
  - Prioritise the same list of high-value datasets adopted by the EU, rather than reinventing the wheel;
  - Review the status of the institutions that govern high-value UK public data, some of which are unsuitable for a modern digital economy.

1. The EU high-value datasets directive

In 2022, the EU required member states to make certain high-value datasets, including data on addresses, buildings, land use and companies, freely available.[1] This is based on a 2019 directive introduced to "fully exploit the potential of public sector information for the European economy and society", particularly SMEs.[2] It focuses on datasets that lack personal information, which can be made freely available without privacy concerns.

The EU aims to support SMEs by reducing data costs and barriers to entry.[3] It also aims to support research and innovation and better-informed policymaking. Previously, charging for data limited innovation, but improving SMEs' access to public-sector information "translates into more jobs and revenue from taxes".[4] The EU consulted on the datasets to include and considered factors including the data's potential to produce value, to benefit many users, and to be combined with other datasets.[5]

The datasets must be made available for free, machine-readable, via APIs and as a bulk download where relevant.[6] Member states have until June 2024 to meet the requirements, and most states are on track to do so.[7] The EU also encourages members to make additional core datasets available; the principle of the directive is that data should be considered open by default (unless it is personal or sensitive), and that systems must build in openness by default at the design stage.[8]

1.1 Initiatives in other countries

Many comparable nations, including some of the UK's key non-EU trading partners, have previously improved access to similar public-sector datasets, with positive impacts. Typically these began as open data initiatives, but have evolved into broader national information infrastructure strategies.

The US

Improving public data has strong political support in the US, and in 2019, the OPEN Government Data Act became law. This requires federal agencies to publish their information online as open data, using standardised, machine-readable data formats, and

---

[1] Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use.
[2] Directive 2019/1024/EU of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.
[3] European Commission, 'High-value datasets: Questions and Answers' (2023).
[4] European Commission, 'Commission Staff Working Document. Impact Assessment Accompanying the document: Proposal for a Directive of the European Parliament and of the Council on the re-use of public sector information' (2018), p.11.
[5] Commission Implementing Regulation (EU) 2023/138 as above.
[6] Directive 2019/1024/EU as above.
[7] European Commission, 'Open Data Maturity' (2022).
[8] Commission Implementing Regulation (EU) 2023/138 as above.

maintain data inventories.[9] The 2021 Federal Data Strategy puts forward an "Open Data Plan" that identifies specific priority data assets and requires the Federal Geographic Data Committee to flag high-value geospatial datasets.[10] The US is now developing an authoritative, open National Address Database (NAD), aggregating state-level address data. [11] In addition, the National Map contains basic open geospatial information, including boundaries, elevation, hydrography, public structures, and land cover.[12]

Australia

Australia released a Public Data Policy Statement in 2015 that committed to make non-sensitive data open by default, requiring public entities to publish data in a machine-readable format, via APIs, alongside relevant metadata; its current national data strategy repeats the commitment.[13] Key datasets have since been identified and prioritised for publication, including health, spatial, transport, environmental, and real-time emergency data.[14] Currently, publicly available data includes an authoritative Australian geocoded address database, national land use data, and topographic datasets.[15]

The Netherlands

The Dutch National Open Data Agenda was published in 2015 and committed government departments to prioritising the release of high-value datasets.[16] This included the development of a centrally-financed system of basic registers, a collection of core datasets including addresses, land, income, and vehicles, which are now available.[17]

The move had positive economic impacts, e.g. Deloitte has estimated that every €1 spent opening up the Netherlands' digital elevation data generated €3.54 in benefits.[18] Previously, an EC study found that after the Netherlands' weather data was opened up in 1999, private-sector reusers generated four times more revenue[19], which contributed an additional €3.5 million per year in corporate tax revenue, seven times the cost of the data provision.[20] The Netherlands' high-value dataset policy continues to be iterated, with the Netherlands'

[9] U.S. Chief Information Officers Council, 'Open Government Data Act (2018)'.
[10] Federal Data Strategy development team, 'Federal Data Strategy 2021 Action Plan', p.11.
[11] U.S. Department of Transportation, 'National Address Database'.
[12] US National Geospatial Program, 'The National Map'.
[13] Australian Government, 'Public Data Policy Statement' (2015); 'Australian Data Strategy' (2022).
[14] Department of Communications and the Arts, 'Open government data and why it matters' (2016).
[15] Data.gov.au, 'Geoscape Geocoded National Address File (G-NAF)' (2016); Australian Bureau of Statistics, 'Towards National Land Account 2021' (2021); Data.gov.au, 'Land use of Australia 2010-11 to 2015-16, 250m'; Smart, 'User guide for Land use of Australia 2010-11' (2016).
[16] Open Government Partnership, 'Netherlands: National Open Data Agenda 2016-18' (2015).
[17] Nederlandse Overheid Referentie Architectuur, 'The Current System of Basic Registers' (2023).
[18] European Commission, 'Impact Assessment study on the list of High Value Datasets to be made available by the Member States under the Open Data Directive' (2020), p.274.
[19] European Commission, 'POPSIS, Pricing of Public Sector Information Study' (2011), p.52.
[20] The United Nations Development Programme, 'The Potential Impact of Open Data in Serbia' (2019), p.98.

2019 data strategy calling for the identification of "necessary improvements in the areas of quality and use of data from the system of key registers".[21]

New Zealand

Improving access to high-value datasets is a central part of New Zealand's national data strategy. It was initially regulated by the New Zealand Data and Information Management Principles of 2011, which mandated that government data be open by default, trusted and authoritative, well-managed, reusable, and reasonably priced.[22] Since then, New Zealand has made available national open datasets of addresses, land use, building outlines, cadastral information and other topographical datasets.[23]

More recently, New Zealand's Data Strategy and Roadmap has identified data gaps systematically across the public sector, enabling the government to invest strategically in data by setting out long-term investment needs across the government data system.[24]

Denmark

In 2012, Denmark established an open system of base registers - known as the Basic Data Programme (BDP) - as part of a national strategy to improve efficiency in the public sector and improve productivity in the private sector. The 10 registers covered real property, geography, addresses, streets and areas, persons, water and climate, and companies.[25] It became mandatory for municipalities to use these registers.[26]

The programme is widely regarded as an economic success. PWC found that the value of geospatial data in Denmark more than doubled after only three years of being opened up, reaching roughly €500 million in 2016.[27] The data has allowed Denmark to develop innovative solutions to problems in both the private and public sectors. For example: drawing on data opened up by the BDP, Denmark was able to develop a Covid vaccination app within a month, and during the recent energy price crisis, data was used to identify gas-heated homes and automatically pay households an appropriate subsidy.[28]

---

[21] The Ministry of the Interior and Kingdom Relations, 'NL DIGITAAL: Data Agenda Government' (2019), p.34.

[22] Data.govt.nz, 'New Zealand Data and Information Management Principles' (2011).

[23] Toitū Te Whenua Land Information New Zealand, 'NZ Addresses' (2021); Ministry for the Environment - Manatū Mō Te Taiao, 'New Zealand land use map' (2023), 'NZ Building Outlines' (2019), 'Browse GIS data'.

[24] New Zealand 'Data Strategy and Roadmap' (2021).

[25] For further detail on the datasets involved, see The Danish Government and Local Government Denmark, 'Good Basic Data for Everyone - A Driver for Growth and Efficiency' (2012); see also EC 'Grunddata (Basic Data)'.

[26] Nicolas Lemcke Horst, Sara Bjerre, Morten Lind, and Line Hvingel, 'The Basic Data Programme - A Danish Infrastructure Model for Public Data' (2014).

[27] PWC, 'The impact of the open geographical data - follow up study' (2017).

[28] Agency for Data Supply and Infrastructure, 'The Danish Basic Data Program, the Data-Distributor (Datafordeleren) and the Covid-19 Pandemic' (2022).

Denmark continues to update its approach to open basic data, releasing a national strategy in 2023 which "includes the next investment in the quality and availability of public data".[29]

Singapore

Singapore launched a national open data portal in 2011, and digital government initiative, Smart Nation, in 2014, which led to a focus on the quality of available public data. The government now ensures key datasets are machine-readable, with standardised formats and metadata.[30] Core information is made freely available on a portal, including land use and occupancy data[31] (with an accompanying statutory land use 'Master Plan'), and cadastral parcels.[32] Building footprint data is seemingly released alongside the Master Plan, and the site also contains open national map data, including road information and contours.

South Korea

South Korea's Act on the Promotion of the Provision and Use of Public Data came into force in 2013. This holds that high-quality and machine-readable public data should be universally accessible, and establishes accountability and governance provisions to ensure as much value is extracted as possible.[33] The National Core Data Portal was subsequently instituted, regulated by triennial data strategies and standards to ensure public data interoperability.[34] The government identifies national core data through feedback from an Open Data Forum.[35] A 2023 initiative continues this momentum, promising to scrap regulations or systems that impede access to key datasets.[36]

Published datasets include procurement data, comprehensive real estate information including land use, construction data, information on environmental impact, and address data via an open API.[37] The data portal also contains land use information, which appears to be linked to other data such as ownership, real estate price, cadastral information, and building information.[38]

2. The UK's approach to high-value data

The UK does not currently have a similar strategy for identifying, or improving the availability of, high-value datasets. It thus appears to be falling behind many trading partners that the UK might reasonably view as comparators on data policy.

---

[29] Agency for Digital Government, 'Data'.

[30] GovTech Singapore, 'Factsheet: Data.gov.sg'.

[31] Data.gov.sg, 'Master Plan 2019 Land Use Layer'; Urban Redevelopment Authority, 'Master Plan'.

[32] Data.gov.sg, 'Collections', 'Master Plan 2019 Building layer', 'National Map Line'.

[33] Korea Law Translation Center, 'Act on Promotion of the Provision and Use of Public Data 2013'.

[34] OECD, 'OECD OURdata Index: 2019'.

[35] Hyejeong Lim, 'Open Data to Improve Transparency and Drive Growth' (2023).

[36] Koh Jean, 'Korea's new innovation strategy: Digital Platform Government' (2023).

[37] Data.go.kr, 'National Core Data: Road name address information'.

[38] Data.go.kr, 'National Core Data: Comprehensive Real Estate Information'.

In the UK, re-use of public sector information is currently covered by the RPSI Regulations 2015, a local implementation of EU legislation.[39] However, unlike the EU and other countries, the UK has not continued to develop a statutory framework for public-sector information, or set out a specific policy for high-value datasets. A government plan to draw up a list of datasets that would form the UK's National Data Infrastructure was abandoned by 2015.[40] Subsequently, access to public data was largely left out of the scope of the National Data Strategy published in 2020, despite the strategy's commitment to "unlock the value of data across the economy".[41]

Although work began in 2022 to create a catalogue of public-sector high-value datasets, this 'Data Marketplace' is currently intended for internal government use only. In 2023, a review by Sir Patrick Vallance warned that the private sector's ability to access data remained "inconsistent and fragmented", and recommended the government should ensure appropriate access to data sets that can "realise commercial value opportunities".[42] In response, the government has agreed to explore how the Data Marketplace could be expanded, but no commitments have been made.[43]

Poor public data has contributed to the UK falling behind its peers in various international comparisons. In 2022, the UN ranked the UK 11th for digital government maturity, down four places from 2020.[44] More recently, the UK has fallen in the international AI Readiness Index, scoring poorly on data availability.[45]

2.1 The state of the UK's high-value datasets

Our research compared the list of high-value datasets identified by the EU with the equivalent datasets in the UK (see the Annex for the full comparison).

We found that several of the UK's equivalent high-value datasets are still only available under complex commercial licences, which limits their value. These licences create legal and administrative costs for businesses that want to re-use high-value datasets, in addition to the licensing costs themselves. Commercial licensing models also make those datasets unusable for some purposes - for example, if a business has an application that requires it to share the data with customers on the public web.

The most important such datasets in the UK are:

1. **Buildings and other topographical information**, managed by Ordnance Survey (OS). Currently OS sells much of its most useful mapping data under complex licence agreements, which limits its usefulness for innovation, especially for data

---

[39] Information Commissioner's Office, 'Guide to RPSI'.
[40] Cabinet Office, 'National Information Infrastructure' (2013).
[41] Department for Science, Innovation and Technology, and Department for Digital, Culture, Media and Sport, 'National Data Strategy' (2019).
[42] Sir Patrick Vallance, 'Pro-innovation Regulation of Technologies Review' (2023), p.9-10.
[43] HM Government, 'HM Government Response to Sir Patrick Vallance's Pro-Innovation Regulation of Technologies Review' (2023).
[44] United Nations, 'e-Government Survey: The Future of Digital Government' (2022).
[45] Oxford Insights, 'Government AI Readiness Index' (2022).

science and research. Some information is now available to SMEs via the 'Geovation' startup programme, but with heavy restrictions on usage.

2. **Land use and agricultural parcels**, managed by the Rural Payments Agency. This is not publicly available, limiting its use for innovation, and for housebuilders and agencies involved in land use planning. The Land Use in England Committee (House of Lords) recommends improving access to land use data, saying "[land use] data must be accessible, understandable and user-friendly at all levels".[46]

3. **Postal addresses**, managed by Royal Mail and OS. These are only available under complex commercial licensing terms, and also create downstream restrictions on the many other other public datasets that contain addresses. (See our full briefing for more information on the UK's postal address data.[47])

4. **Meteorological information**, managed by the Met Office. While some of the Met Office's weather data is available openly, much of it is sold on a commercial basis. In 2015 the BBC stopped using Met Office data in its weather forecasts for cost reasons.[48]

See the Annex for a full analysis of the UK status of the set of EU high-value datasets.

2.2 The governance of the UK's high-value data

The underlying cause of many of these issues is the ownership and governance situation of these datasets.

For example, ownership of intellectual property (IP) rights in the UK's national address database creates difficulties for the use of address data and other public-sector datasets that contain addresses. The most widely used address data product is the Postcode Address File (PAF) maintained by Royal Mail, a private company.[49] Companies that wish to access this must pay a fee and sign legal agreements. In addition, several important public datasets, such as information on property transactions and on the energy performance of buildings, are not fully open because they contain address data from PAF that is only available on restricted terms.[50]

More generally, some of the UK's most valuable public data, such as data on buildings and weather, is controlled by 'government-owned companies' or trading funds, including the Ordnance Survey and Met Office. Such organisations are required to operate as commercial enterprises maximising the immediate returns from their data assets, which generally means selling expensive licences to large incumbents, rather than considering the broader economic value of their data.

---

[46] The House of Lords Land Use in England Committee, 'Making the most out of England's land' (2022), p.70.
[47] Centre for Public Data, 'Why the government should open up the UK's address data' (2023).
[48] BBC, 'Met Office loses BBC weather forecasting contract' (2015).
[49] Royal Mail, 'PAF®'.
[50] HM Land Registry, 'HM Land Registry: Price Paid Data'; Department for Levelling Up, Housing and Communities, 'Energy Performance of Buildings Data: England and Wales'.

Other important datasets are controlled by arms-length bodies, including the Rural Payments Agency, Environment Agency, and Valuation Office Agency, which maintains data on business valuations.

To address these issues, the government should review whether it is appropriate to run strategic data institutions in this way. HM Land Registry may provide a precedent here: it became a trading fund in the 1990s, but this ended in 2020.[51] Other countries have reformed the governance of such institutions: for example, the Netherlands re-nationalised its meteorological institute in 1999.[52]

3. Recommendations

The UK is falling behind other countries in making its high-value data available to the private sector and civil society. This is partly a consequence of not having adopted a national strategy on access to high-value data, and partly a consequence of some of its high-value datasets being managed by trading funds, government-owned companies and arms-length bodies.

This difference in data policy may increase business costs and harm the UK's international reputation. Access to high-value data in the EU will raise the expectations of SMEs and other businesses and increasingly, companies trading in the UK will be disappointed if they do not find the same standard of public data infrastructure. The UK will also suffer opportunity costs if it does not open up key datasets, as startups and SMEs will find it harder to innovate, and lower data quality will hamper public service delivery.

To tackle this, we recommend the government should:

- Adopt a national high-value data strategy as part of its wider industrial strategy. This should identify initial datasets to prioritise for improvement; create a clear strategy for identifying further high-value datasets, which should include consultation with the private sector and civil society; and address any governance changes required. New Zealand, the Netherlands and Denmark all provide examples of such strategies.
- Use the EC's list as a basis for selecting high-value datasets to prioritise. Our international comparative research suggests that the set is fairly standard across countries. The UK can run additional consultation on UK-specific datasets that could help unlock value, but there is no need to reinvent the wheel.
- Mandate that high-value datasets be made available under open licences, in a machine-readable format, and via both API and bulk download.
- Prioritise improving access to the following key high-value datasets in the UK:
    - postal addresses
    - buildings and topographical information

---

[51] HM Land Registry, 'Explanatory Memorandum to the Land Registry Trading Fund (Revocation) Order 2020' (2020).
[52] Peter Weiss, 'Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts' in Fitzgerald (ed.) 'Access to public sector information: law, technology and policy: Volume 2' (2010).

○ land use and agricultural parcels
○ meteorological information.
● Review the governance of certain key data-producing institutions, to ensure the UK remains competitive with other countries.

Annex: Full list of EU 'high-value datasets' and their UK status

This annex details the major high-value datasets listed in the EU Implementing Regulation and analyses their current status in the UK. We highlight in red where the UK is lagging.

● Where data is only available for some of the UK's devolved nations, we note this.
● We exclude the 'statistics' theme, which is generally covered by earlier directives and where the UK tends to have good information.
● We exclude also the 'mobility' and 'earth observation and environment' themes, which are covered by an earlier EU directive that is still in force in the UK, and where data has been made broadly available. Land use data, covered by the earth observation theme, is one exception to this rule, and is therefore included.

Theme and dataset	Required attributes	UK status
Geospatial - Administrative units	Unique identifier; Unit type (administrative or maritime unit); Geometry; Boundary status; National identification code; Identification code of the upper administrative level; Official name; Country code; Name.	Ordnance Survey's Boundary-Line dataset (GB) is published under OGL[53] and has many administrative boundaries. ONS also supplies various administrative geography datasets.
Geospatial - Geographical names	Unique identifier; Geometry; Name; Type.	Open via Ordnance Survey's OpenNames dataset (GB), which is published under OGL.
Geospatial - Addresses	Unique identifier; Geometry; Address locator (e.g. house number); Thoroughfare (street); name; Administrative units (e.g. municipality, province, country; Postal descriptor (e.g. post code); Date of last update.	Users must pay and sign licensing agreements for either: ● Royal Mail's Postcode Address File, which contains addresses and identifiers ● OS's AddressBase (GB), which contains addresses, identifiers and locations.

---

[53] The Open Government Licence (OGL) is the default licence for open data published by the UK public sector.

		OS's OpenUPRN (GB) contains UPRN, aka postal delivery points, with associated locations, but UPRNs are not one-to-one with buildings. In addition, many other public-sector datasets also contain addresses (for example, energy performance data), and downstream IP complexity also affects those datasets.
Geospatial - Buildings	Unique identifier; Geometry (footprint of the building); Number of floors; Type of use.	OS's Mastermap Topography Layer (GB) contains buildings (and much more besides) but is not open. A government plan to release an open dataset of property extents created from the OS MasterMap Topography Layer was announced in 2018 but subsequently abandoned. Additionally, data 'derived' from OS topography info (e.g. drawn on OS maps) has historically been held to contain OS IP. Consequently, downstream IP issues have affected many other valuable public-sector datasets - for example, rights of way data managed by local authorities. The 'presumption to publish' approach should help, but is cumbersome to use in practice. A new National Buildings Database (GB) is being developed, but the licensing of this data is not yet clear.
Geospatial - Cadastral parcels	Unique identifier; Geometry; Parcel or basic property unit code; A reference to the administrative unit of lowest administrative level to which this parcel or basic property unit belongs.	Theoretically open, but not usefully. The INSPIRE Index Polygons are cadastral parcels for England and Wales published under an open licence, but in a way that can't be joined to title numbers. They also only cover freehold titles.

		The INSPIRE Index Polygons are a subset of the National Polygon Dataset, which is not freely available and contains more authoritative attributes (such as the title number).
Geospatial - Reference parcels	Unique identifier; Geometry (boundary and area); Land cover; organic); Stable landscape elements ("EFA-layer"); areas with natural/specific constraints.	Land parcels are held by the Rural Payments Agency in the Rural Land Register, but this data is not openly available. The UKCEH Land Cover Maps (UK) dataset is available for re-use, but not on open terms.
Geospatial - Agricultural parcels and their land use	Unique identifier; Geometry (boundary and area of each agricultural parcel); Land uses (crops or crop groups); Organic; Individual landscape element; Permanent grassland.	As above. Some land use statistics for England are available under Land Use Statistics, but not to the required level of granularity.
Earth observation and environment - Land use	Territory characterised according to its current and future planned functional dimension or socio-economic purpose (e.g. residential, industrial, commercial, agricultural, forestry, recreational).	None apparently open. The Land Use Statistics for England provide a breakdown with the required granularity of land use classifications, but only break down to local authority level. The National Land Data Programme ran some standalone projects to improve the quality of some of the data, but there appear to be no plans to roll this out further.
Meteorological - Observations data measured by weather stations	Per weather station, full temporal resolution; All observation variables measured	Contained in MIDAS-Open (UK), held by CEDA. Requires registration, but licensed under OGL.
Meteorological - Climate data: validated observations	Per weather station, full temporal resolution; All validated measured observation variables; daily average per variable	As above, via MIDAS-Open.
Meteorological - Weather alerts	Alerts, 48 hrs or more ahead	Unclear, but apparently not available as open data.

Meteorological - Radar data	Per radar station in the MS and national composite; Reflectivity, Backscatter, polarization. Precipitation, wind, and echotops	Rain radar data is available from the Met Office's NIMROD system (UK), but this is not open.
Meteorological - Numerical weather prediction model data	Minimum 48 hrs ahead in 1hr steps, national, at 2.5km/best available grid; Deterministic and/or ensembles if available, for meteorologically relevant parameters and levels	Unclear, but apparently not available as open data.
Companies and company ownership - Basic company information: key attributes	Name of the company; Company status; Registration date; Registered office address; Legal form; Registration number; Member State where the company is registered; Activity/activities that are the object of the company.	This is open for practical purposes - see the Companies House Free Company Data product (UK). This is not explicitly openly licensed, but is made available without restriction.
Companies and company ownership - Company documents and accounts	Accounting documents, which include: Financial statements (incl. the list of participating interests, subsidiary undertakings and associated undertakings, their registered office address and proportion of capital held), audit reports; Non-financial statements, management reports and other statements or reports; Annual financial reports.	This is partly open - see the Companies House Free Accounts Data Product. Licensing as above.

About and acknowledgements

The Centre for Public Data is a non-partisan, non-profit research and advocacy organisation that works to improve the quality of UK public data. We are technologists who work with Parliamentarians, policymakers, civil servants and civil society to improve data-related provisions in policy, legislation and regulation.

This briefing was written by Anna Powell-Smith and Amber Dellar at the Centre for Public Data. Thanks to Peter Wells and Owen Boswarva for comments; any errors are our own.