

ENERGY RELATED GIS DATA LANDSCAPE

DAMON ROBERTS

DIGITAL AND DATA CONSULTANT

25 MARCH 2022



CONTENTS

1.	Introduction	1
2.	Ordnance Survey.....	3
2.1.	OS Open Data	3
2.2.	OS Closed Data.....	6
3.	OpenStreetMap.....	7
3.1.	How to use OSM.....	7
3.1.1.	Extracting Data.....	7
4.	Other Commercial GIS data providers.....	9
5.	Adjacent Datasets.....	10
5.1.	Electricity Network Data.....	10
5.1.1.	Transmission System Operator – National Grid Energy System Operator (ESO) 10	
5.1.2.	Distribution Network Operators (DNO).....	10
5.2.	Gas Network Data.....	10
5.1.	Elexon.....	10
5.2.	XOServe.....	10
5.3.	Smart Meter Data - Smart Data Communications Company (DCC).....	10
5.4.	Power Generation Data	11
5.4.1.	Sheffield Solar.....	11
5.4.2.	Renewable UK.....	11
5.5.	Data.gov.uk	11
5.6.	Office for National Statistics (ONS) & Census Data.....	11
5.6.1.	Census Data Terminology and Use	12
5.7.	UK Country-Specific Datasets	12
5.8.	Emissions Data	13
5.9.	Energy Performance of Buildings Data.....	13
5.10.	Onshore Geological Data.....	13
5.10.1.	British Geological Society (BGS)	13
5.10.2.	UK Onshore Geophysical Library.....	13
5.11.	Offshore Data.....	13
5.11.1.	Oil & Gas Authority National Data Repository	13
5.11.2.	ADMIRALITY Dataset	14
5.11.3.	British Oceanographic Data Centre (BODC).....	14
5.11.4.	Aeon Public Data Repository (PDR)	14

5.11.5. Marine Data Exchange	14
5.12. Open Data Collections	14
5.12.1. ESRI	14
5.12.2. Soar.Earth	14
5.12.3. Natural Environmental Research Council (NERC) Data Catalogue.....	15
5.13. GIS Dataset Listings (Mixed Licenses)	15
5.14. Regen Local Energy Data Innovation Report	15
6. Commercial Datasets.....	16
Appendix – GIS Data Formats.....	17

DISCLAIMER

This document has been prepared by Energy Systems Catapult Limited. For full copyright, legal information and defined terms, please refer to the "Licence / Disclaimer" section at the back of this document.

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. INTRODUCTION

The landscape for geographic information system (GIS) and other locational datasets for Great Britain (GB) is a particularly busy area, with a vast range of suppliers, data types and licence terms. Often, the most detailed data can come with complex terms of use, and a prohibitive cost. The scope of this report only covers GB, as Ordnance Survey (OS) is split down into OS, OS Northern Ireland, and OS Ireland, and the latter two have not been covered. Other data sources mentioned may include data for Ireland and Northern Ireland, as well as the rest of the British Isles and beyond.

This report details some of that knowledge and aims to give an insight into the landscape for GIS and other locational datasets, focussed on GB. There is a massive amount of data out there, but much of it lacks discoverability and often one has to spend a long time searching through data portals and other resources to find what they are looking for.

The report begins with an overview of the most common GIS data formats. Following, section 2 gives information on open and closed data available from OS, as they are the seminal source for cartographic data in the GB. Section 3 explores similar or equivalent datasets available from other sources, such as Open Street Map (OSM).

Later sections go on to explore data that can be used alongside pure locational data, such as governmental datasets, census results, marketing demographics, and private/commercial data. It is by no means a complete picture and does not aim to be an exhaustive list of data sources, however it is comprehensive enough to give a grounding on where to start looking, and what to look for.

Table 1 lists the data sources covered by the report, and links to the appropriate section.

Table 1 – Data Categories

Data Type	Description	Shortcut(s)
GIS data providers		
Mapping/Geographical	Listings and comparison of OS and Non-OS data sources. Correct as of 12/2021	<ul style="list-style-type: none"> • Ordnance Survey • OpenStreetMap
Adjacent data providers		
Energy	Energy specific datasets & data providers.	<ul style="list-style-type: none"> • Electricity Network Data • Gas Network Data • Elexon • XOServe • Smart Meter Data - Smart Data Communications Company (DCC) • Power Generation Data
Government/Varied	Wide range of open and closed datasets collected or provided by governmental or its agencies.	<ul style="list-style-type: none"> • Data.gov.uk • Office for National Statistics (ONS) & Census Data • UK Country-Specific Datasets <p>The Welsh Government maintains their own geoportal, collating a wide range of datasets, including datasets related to planning wind power, climate change, and air quality. Much of the data is under an open government licence making it accessible to innovators.</p> <p>Similarly, the Scottish government has their own INSPIRE formatted data portal, covering a very similar remit to the UK's.</p> <ul style="list-style-type: none"> • Emissions Data • Energy Performance of Buildings Data • UK Country-Specific Datasets British Geological Society
Geological	Includes onshore surface and subsurface data, as well as offshore.	<ul style="list-style-type: none"> • British Geological Society • UK Onshore Geophysical Library • Offshore Data

<p>Scientific/Varied</p>	<p>A wide variety of datasets and data repositories from various sources, both commercial and open. Covers all of the above categories, and others.</p>	<ul style="list-style-type: none"> • Open Data Collections • Natural Environmental Research Council (NERC) Data Catalogue <p>The NERC data catalogue is a portal that hosts a broad range of research related datasets, spanning earth, marine, polar, and atmospheric science, many of which are UK centric. Datasets span from historic (1800s) to the present day. Some specific topics that may be of use are changing land use and emissions data.</p> <ul style="list-style-type: none"> • GIS Dataset Listings (Mixed Licenses) • Commercial Datasets
--------------------------	---	--

2. ORDNANCE SURVEY

Ordnance Survey is the definitive source of GB mapping data. Recently, they have begun offering a number of datasets under the [open government licence \(OGL\)](#), however their most detailed data, the parts typically required by innovators and local scale modelling teams are still premium, copywritten products only available commercially.

During development, data is offered [at no cost to innovators](#). In some cases, a local authority or other government entity, by having a [Public Sector Geospatial Agreement \(PSGA\)](#), can pass data onto an end user through a [Public sector licences for contractors](#), but for commercial uses the users will still need to form a partnership agreement with OS.

Below are two tables of OS Datasets; the first is their open data, which are free to use under an OGL style licence, as long as OS is attributed. The second is premium, paid data. Where equivalent datasets from other suppliers have been identified, they are detailed alongside.

2.1. OS OPEN DATA

Table 2 gives information on OS Datasets which are free to use under an OGL style licence, as long as OS is attributed. Where equivalent datasets from other suppliers have been identified, they are detailed alongside.

Table 2 – OS Open Datasets

OS Open Dataset	Description	Filetype	OSM Equivalent/Other Info
1:250 000 Scale Colour Raster™	Roadmap style map of GB.	Tiff	OSM does not provide raster images by default, and does not recommend creating them unless they are needed for lightweight visualisation purposes, and lists several tools for doing so. OSM's default 'unit' for breaking a map down is in tiles, the tile page on the wiki goes into more detail.
Boundary-Line™	Outlines of council wards/administrative areas.	Multiple GIS formats	OSM-boundaries offers a GUI tool for extracting administrative boundary data in OSM, and covers most of the world in varying levels of detail.
Code-Point® Open	Centrepoin'ts of postcode areas, but not polygons	Multiple GIS formats	OSM provides a tool that details the state of GB postcode area mapping. The rest of the site has some useful stats and info for newcomers to OSM.
GB Overview Maps	Simple GB map	Geotiff	Many suppliers listed on the OSM Wiki (detailed in How to use OSM) offer similar maps.

OS Open Linked Identifiers	A collection of OS location identifiers, used to tie their various products together.	CSV.	These datasets enable linking of different OS datasets via unique identifiers. They do not contain location data, however could potentially be used as a 'key' to link OS data to other (such as OSM) datasets grouped by street, area etc.
MiniScale®	Simple image-based map of GB.	EPS, TIFF and other raster image formats.	Many suppliers listed on the OSM Wiki (detailed in How to use OSM) offer similar maps.
OS Open Greenspace	Covering a range of greenspaces in urban and rural areas including playing fields, sports' facilities, play areas and allotments.	Multiple GIS formats	This page on the OSM wiki details the various green space tags used in OSM. Currently, there are no public sources for this data directly, and therefore extracts either need to be generated manually using Overpass (detailed in How to use OSM) or a 3 rd party OSM supplier.
OS OpenMap - Local	Standard street maps.	Multiple GIS formats	Readily available directly on OSM and various 3 rd party sources based on OSM data.
OS Open Names	Placenames and Postcodes.	CSV, GML & Geopackage	Readily available directly on OSM and various 3 rd party sources based on OSM data.
OS Open Rivers	Understand how watercourses in Great Britain join up.	ESRI® Shapefile, GML, GeoPackage, and Vector Tiles	OSM holds a range of data about GB waterways, detailed on their Wiki . There is an open-source project available on Github which extracts all river related data, however it appears to no longer be supported by the developer.
OS Open Roads	A high-level view of the road network, from motorways to country lanes.	ESRI® Shapefile, GML, GeoPackage, and Vector Tiles	Readily available directly on OSM and various 3 rd party sources based on OSM data. Tags and details are on the OSM Wiki Highways page .

OS Open TOID	<p>A TOID (Topographic Identifier) is a unique and persistent identifier for each and every feature found in OS MasterMap products. Open TOID is an open dataset providing access to a generalised location of key features found in OS MasterMap, enabling visualisation of third-party data linked to their respective TOID identifier.</p>	<p>CSV, and GeoPackage</p>	<p>TOIDs are not included in OSM. They are mainly used as a 'glossary' for what OS dataset covers a particular feature.</p>
OS Open UPRN	<p>An open dataset containing all the Unique Property Reference Numbers (UPRNs) found in AddressBase Premium, with their respective geometries in British National Grid and Latitude, Longitude.</p>	<p>CSV, and GeoPackage</p>	<p>UPRN's are listed on OSM, however, they do not offer a way to tie them to specific street addresses as this information has not been released with a suitable licence.</p>
OS Open USRN	<p>An open dataset of all Unique Street Reference Numbers (USRNs) within OS MasterMap Highways Network, with an associated simplified line geometry representing the geographic extent of each USRN.</p>	<p>GeoPackage</p>	<p>USRN's are fully supported and mapped within OSM.</p>
Strategi®	<p>Deprecated detailed map. Last updated 2016.</p>	<p>DXF, ESRI® Shapefile, and MapInfo® TAB</p>	
OS Terrain® 50	<p>3D Landscapes</p>	<p>ASCII Grid and GML (Grid), ESRI® Shapefile, GML, GeoPackage, and Vector Tiles</p>	<p>By default, OSM does not include any height mapping data, as it is extremely computationally heavy, and GPS devices do not provide very good vertical data, and therefore goes against the Crowdsourced ethos of OSM. The Srtm2Osm tool can generate contours that can be added to an OSM file.</p>
OS VectorMap® District	<p>This is a service offered by OS to download sections of their map as vector files, with standard Streetmap data.</p>	<p>ESRI® Shapefile, GML, GeoPackage, and GeoTIFF</p>	

2.2. OS CLOSED DATA

Table 3 contains OS closed datasets. These are typically an enhanced version of their open ones, either with more detail in the form of polygons, or more metadata allowing easier linking between objects.

Table 3 – OS Closed Datasets

OS Closed Dataset	Description	OSM Equivalent/Other Info
OS MasterMap Topography Layer	The most detailed and accurate GB map source.	OSM aims to replicate this dataset. The combination of all OSM layers will likely have equivalent or greater amounts of information in some areas but will have gaps in others.
AddressBase	Postcodes, co-ordinates, and classifications. Has various tiers with different levels of access and details.	Google Maps places API can offer similar details, and OSM includes some but not all address data in a similar form.
OS VectorMap Local	A highly detailed, customisable, street-level map, OS VectorMap® Local shows fences, building outlines, paths and street names, for the accurate lie of the land.	OSM aims to replicate this dataset but currently is not as comprehensive as OS.
Code-Point with Polygons	Code-Point® with Polygons shows the notional shape of every postcode unit in Great Britain and includes major buildings with multiple postcodes.	This tool details the state of GB postcode area mapping on OSM. The rest of the site has some useful stats and info for newcomers to OSM.
Points of Interest	Points of Interest is the most comprehensive, location-based directory of all public and privately-owned businesses, education and leisure services in Britain.	Points of interest are a focus of many mapping tools, and often utilize crowdsourced data. Google Places API, OSM and Bing all rely on user submitted data for their 'place' data, but have differing levels of verification and oversight. OSM is the only provider to release the data openly.
AddressBase Plus	Along with Royal Mail PAF data, AddressBase® Plus gives you up-to-date local authority addresses, multi-occupancy addresses and OS MasterMap TOID references.	This is an area where OS is very difficult to compete with. They have details of 30 million plus addresses in GB, whereas OSM has fewer than 3 million.
OS MasterMap Greenspace layer	Shows accessible and non-accessible greenspaces in urban areas	This page on the OSM wiki details the various green space tags used in OSM. Extracts either need to be generated manually using Overpass (detailed in How to use OSM) or a 3 rd party OSM supplier like GeoFabrik or MapBox.
OS Features API	Combines free and paid data into an API, combining land use data and location data into a single request.	Similar to the OSM Overpass API. This level of access is good while experimenting with models and solutions in a small area.

3. OPENSTREETMAP

Open Streetmap is not a direct competitor to OS, as they do not have any direct commercial goals, however, they do aim to achieve a dataset equal to, or exceeding OS.

OSM began as a project aiming to crowdsource mapping data, generating it within the public domain and to compete directly with OS, who OSM state to have a government sponsored monopoly in GB mapping data. A key principle that OSM follows is that all content needs to be user generated. Several fully open datasets, in the GB, those released under the open government licenced have been added, but by and large everything is user drawn. Therefore, there are significant gaps and inaccuracies in OSM data, particularly in less populated areas. Urban centres tend to be very well mapped, there is little to no evidence of people maliciously tampering with the data.

OSM data is arranged as a collection of points, shapes, and tags, detailed on the project [wiki](#). Interestingly, many of their datapoints are not directly visible on the web and need to be extracted specifically. This makes identifying what is available time-consuming.

3.1. HOW TO USE OSM

With a view for all these datasets to be used for machine learning, this section will mainly be a primer on how to extract data from OSM, whilst touching on how the OSM editor currently works.

3.1.1. EXTRACTING DATA

The main way to access data from OSM is via their [Overpass API](#). The API runs server side, and therefore instances are hosted by a third party. The API has a soft limit of returning around 10 million map elements, which should be sufficient for most projects, but the API hosts have their own request/rate/data transfer limits. One can also host their own [OSM API server](#), with a live copy of the entire OSM database fitting easily onto a 1tb drive. To experiment with the API query language, there is the [Overpass Turbo tool](#), which is a web-based GUI for Overpass that also returns a graphical representation of the query output. For more details on how to use OSM, both contributing and extracting data, [learnosm.org](#) has a wide range of tutorials and guides. To download data in a more 'archival' format, or for single downloads of the entire OSM 'world' [direct downloads, torrents, and premade docker environments are available](#). [TagInfo](#) provides listings of tags, and their frequency, similar data, and more, and is useful for building queries or looking for tags that may contain required data.

Figure 1 shows an uncleaned extract of OSM 'power' tagged data, comparing its mapping of High Voltage powerlines against OS Open Data for Coventry and its surrounds. The blue is OSM, and the red OS, the 'glowing' purple shows overlaps. OSM includes some underground cabling, and other pieces of data that OS is missing, whilst OS has better coverage of the main conductors. However, some of the small blue spots are substations, incorrectly tagged as Low Voltage conductors on OSM, highlighting the need to manually inspect data sourced from OSM before use.

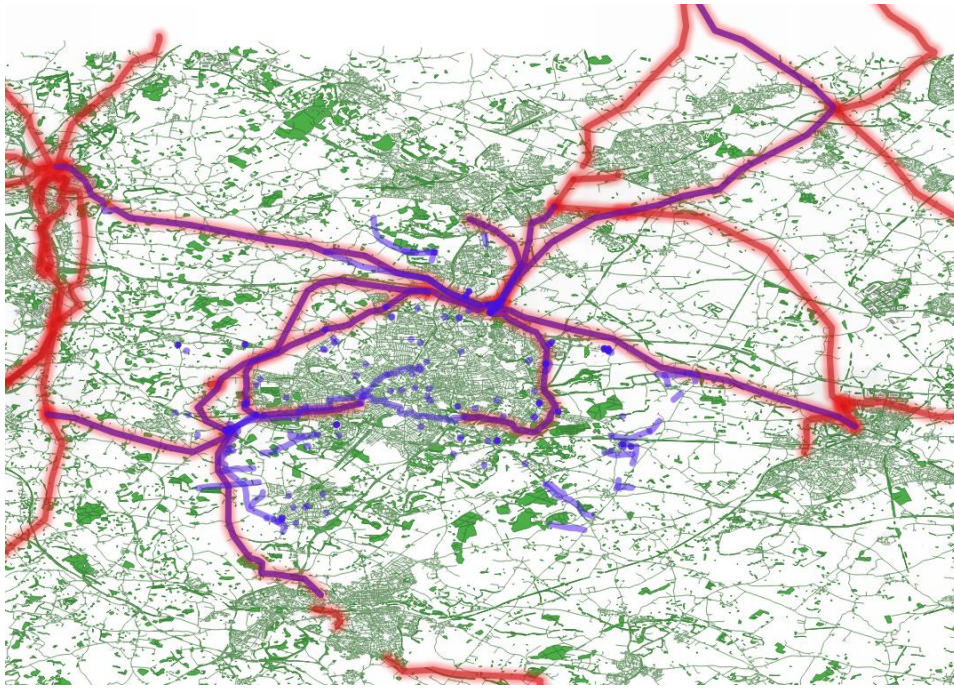


Figure 1 – An extract from a GIS tool comparing OS and OSM.

Various suppliers provide pre-built datasets and offer consultancy style services to aid people in extracting the data they want from OSM. As of the time of writing, these services vary in price, but are normally a one of fee for a particular data extract. Some now offer to build APIs as a service for data that needs more regular updates. A list of suppliers is maintained on the [OSM Wiki](#).

4. OTHER COMMERCIAL GIS DATA PROVIDERS

On the commercial side, Google maps and Bing maps are main alternatives to OS and OSM. Bing are mainly focussed on fleet and routing services, providing accurate, routable basemaps. They have partnered with OSM by contributing machine generated data, such as building shapes and satellite photography for OSM users to trace out features but do not offer much data openly.

Google offers a wide range of API accessible data within the Maps Platform, as well as AI and ML services used to extract data from Streetview and other adjacent datasets they hold. The Places API offers a wide range of functions, some of which are similar to the APIs available from OS, but with a mixture of Google generated, business and owner submitted, and crowdsourced data from Google Maps.

5. ADJACENT DATASETS

5.1. ELECTRICITY NETWORK DATA

5.1.1. TRANSMISSION SYSTEM OPERATOR – NATIONAL GRID ENERGY SYSTEM OPERATOR (ESO)

The ESO operates the UK's high voltage transmission networks and hold full or partial authority over system wide operation, energy trading, demand and constraint management, and future planning. They hold a significant amount of data, available through their [own data portal](#). Most datasets are available via API access and are licenced as open data.

5.1.2. DISTRIBUTION NETWORK OPERATORS (DNO)

The low voltage side of the UK's energy system is operated by 6 major DNOs, with several smaller entities (known as independent DNO's) covering more bespoke connection strategies to the Grid. The data available from DNO's varies dramatically. [WPD](#) and [UKPN](#) offer comprehensive data portals, whilst others share data upon request. The [Energy Networks Association](#) maintains a data request tool, enabling them to broker data requests, and simplify the process of requesting data from each DNO. Data on primary substations can be found in the data tables published as part of the Long-Term Development Statements.

5.2. GAS NETWORK DATA

National Grid also operates the main UK gas distribution network, with gas distribution network (GDN) operators sitting below them, delivering gas to end users. This is very similar to the relationship between the ESO and DNO's. They maintain the [Prevailing View](#) data portal, which offers mainly information about the demands for, and composition of gas in the network. The ENA portal can be used to request gas network data that is unavailable.

5.1. ELEXON

Elexon operate the balancing and settlement mechanism for the UK's electricity network. They publish a wide range of datasets including energy usage and generation composition, spot pricing, system and operational data, balancing services contracting and pricing, amongst others. Most data are available via well documented APIs once registered.

5.2. XOSERVE

XOServe is the main data processor for the UK gas industry, and is responsible for processing invoices for the industry, maintaining the database of gas meters, and operating the balancing and settlement market. They hold a huge amount of data and working on enabling more advanced access to their data through their sister company Correla. At the moment, any requests for data from them are handled on a case-by-case basis and mainly targeted toward GDNs and customer facing services.

5.3. SMART METER DATA - SMART DATA COMMUNICATIONS COMPANY (DCC)

Smart meter data is poised to become a significant asset within energy innovation, however, at the moment data availability is limited. Several dated, static datasets are available, but real-time data is expensive and difficult to access, particularly when considering GDPR requirements adding a layer of complexity in linking consumption to geographical data at a high resolution. The DCC is the

operator of the current smart meter network in the UK, and makes data available either through direct partnership, or through a number of third parties. More information on how to access Smart Meter data can be found [here](#).

5.4. POWER GENERATION DATA

Other than that which is reported by Elexon, the two data sources below offer greater insight into solar and wind generation in the UK.

5.4.1. SHEFFIELD SOLAR

[Sheffield Solar](#) is a project being run by the University of Sheffield, in collaboration with National Grid ESO. Their current aims are to: further increase data resolution through 5 minutely outturn estimates in near-real-time; improve the accuracy of the regional PV outturn estimates through better mapping of the electricity network; and to develop a methodology for estimating the deployed PV capacity using metered demand data and modelled PV yield. Data is available via a set of APIs.

5.4.2. RENEWABLE UK

[Renewable UK](#) contains a significant amount of data on windfarm installations > 100kW. Their data is partially open; however some requires a membership to access, starting at £2150 + VAT at time of writing. Their datasets include capacities and specifications, project statuses, and mapping tools. They also have a smaller wave and tidal power dataset available, but state that it is not comprehensive and is actively being developed.

5.5. DATA.GOV.UK

[Data.gov](#) is one of several central data repositories/catalogues run by the UK government, this is the most complete and covers an extremely wide range of data. Many energy related datasets are listed, a few examples being council EV charger use, household insulation levels, and emissions data. More broadly, it lists both open, and closed datasets, spanning from food safety to crime rates and government spending. Access to closed data can be requested, either purely on the merits of the request, or under a [Freedom Of Information \(FOI\) request](#). Under the [EU INSPIRE](#) directive, all EU member states agreed to a common data framework for Geospatial data. This directive led to the creation of the [Geospatial Commission](#), who have recently (November 2020) published a [comprehensive market study and action plan](#) for the next few years. A critique and review of their approach and progress, with a view toward open data is currently maintained by [Owen Boswarva](#) and has been since 2017. Alongside this, the Geospatial commission also maintains a list of [data 'themes'](#) and Government linked bodies that hold data matched to that topic.

5.6. OFFICE FOR NATIONAL STATISTICS (ONS) & CENSUS DATA

UK census data, and more widely the [ONS](#) is a valuable, and open source of demographic data. Their datasets are typically OGL licenced, and well-structured .csv files. Whilst often not directly applicable, insights about communities, businesses, and market desires can be derived from their reports. The 2021 UK Census results are due to be reported in Spring 2022, current data is from 2011, and therefore somewhat outdated.

5.6.1. CENSUS DATA TERMINOLOGY AND USE

[Nomis](#) is the portal the ONS uses to share all the data they hold, both from the census, and other studies they have been commissioned to carry out. Aside from census data, the ONS provides a subset of [geographical data](#), most of which are polygons of their various subdivisions, however they also provide some local authority data and outlines.

The census, and other ONS data often use their own system of geographic identifiers to break the country down into small subsections with relatively even populations. These identifiers can then also be used to tie results within an area to other datasets, allowing relatively complex analysis to be carried out with high precision. Full details are on the [ONS website](#) and summarised below. Other historical subdivisions also exist but will not be covered.

BUAs and BUASDs: Built-Up Areas (BUAs) and Built-Up Area Sub-Divisions (BUASDs) are geographical datasets generated by ONS, intended to create a basis for comparisons between built up areas and more rural locales. Some BUA's are listed as not having a population, as they do not contain any residential buildings.

OA: Output Areas (OAs) are the lowest geographical boundaries released by the ONS. They are loosely based on postcode areas. They were designed to have similar population sizes and be as socially homogenous as possible based on tenure of household and dwelling type (homogeneity was not used as a factor in Scotland). They attempt to form sensible shapes and tend to be bounded by roads or other similar features where possible. The minimum OA size is 40 resident households and 100 resident people, but the recommended size is larger at 125 households. Larger groups of OA's exist, called lower layer super output areas (LSOAs), middle layer super output areas (MSOAs), and upper layer super output area (USOAs). USOAs have not yet been defined in the UK. Each layer is built up from the layer below it, so bounds are common between them. These bounds do not apply to Scotland or Northern Ireland.

Table 4 – Output Area Definitions

Area type	Lower threshold		Upper threshold	
	People	Households	People	Households
Output Areas	100	40	625	250
Lower Layer Super Output Areas	1,000	400	3,000	1,200
Middle Layer Super Output Areas	5,000	2,000	15,000	6,000

It is important to note that OAs are redefined across censuses, so historical data needs care if being brought into the model.

5.7. UK COUNTRY-SPECIFIC DATASETS

The Welsh Government maintains their own [geoportal](#), collating a wide range of datasets, including datasets related to planning wind power, climate change, and air quality. Much of the data is under an open government licence making it accessible to innovators.

Similarly, the Scottish government has their own [INSPIRE](#) formatted data portal, covering a very similar remit to the UK's.

5.8. EMISSIONS DATA

Emissions data is a difficult area, as standards for emissions reporting vary across sectors and business sizes. There are various low-resolution national datasets, and often councils have their own emissions tracking data, that may or may not be open. The central UK source for local air quality datasets is held by [DEFRA](#). This datasets holds two main types of air quality measurement, diffusion tubes, and the [Automatic Urban and Rural Network \(AURN\)](#). It is important to note that diffusion tube measurements cannot really be compared with AURN data, as the AURN takes continuous measurements, whilst the tubes average measurements over their test periods. Often, a city may only have one air quality meter, limiting the usefulness of the data. If accurate air quality data is required, a city-wide diffusion tube-based study over the course of a few months is likely to be required.

Complimenting the local datasets is the [NAEI](#), which offers a county/regional datasets including CO₂ emissions, road transport fuel consumption and non-gas, non-electricity, and non-road transport fuel consumption. These datasets can be linked and compared to worldwide data, as they are typically calculated aggregates rather than direct measurements. Scotland has other environmental datasets available, from the [Scottish Environmental Protection Agency](#).

5.9. ENERGY PERFORMANCE OF BUILDINGS DATA

[Energy Performance Certificate \(EPC\)](#) data covers England and Wales, with data for [Scotland on its own portal](#). EPC provides a measure for how energy efficient a property is. Metadata available from EPC records can also provide a measure of how well insulated houses in an area, as well as information on property types and floor areas. This data can be used for scenario modelling, and also to bolster assumptions made based on demographic data. EPC data does not exist for properties that have not been sold or let since the 1st of October 2008. Moreover, EPCs are only available for privately-owned buildings. Equivalent information for public buildings is available through Display Energy Certificates (DECs) data.

5.10. ONSHORE GEOLOGICAL DATA

5.10.1. BRITISH GEOLOGICAL SOCIETY (BGS)

[The BGS](#) holds a significant amount of data about the UK. The majority is subsurface data; however they also have a wide range of historical and photographic information. Many of their datasets are available via APIs, and BGS also provide various pieces of software to use alongside their datasets.

5.10.2. UK ONSHORE GEOPHYSICAL LIBRARY

[The UK Onshore Geophysical Library \(UKOGL\)](#) manages the archive and official release of 2D and 3D seismic data recorded from inland areas of the UK. It is a self-sustaining independent charity which receives no funding from Government.

They work with the [Oil & Gas Authority](#) (OGA) with the continued long-term objective of bringing all available UK onshore seismic data into secure archival storage, while providing support, information and open access to its archives for all interested parties.

5.11. OFFSHORE DATA

5.11.1. OIL & GAS AUTHORITY NATIONAL DATA REPOSITORY

The [National Data Repository](#) (NDR) is operated by the OGA. It aims to be a single source of information for all open data within Oil and Gas industries, both onshore and offshore. At the

moment, it mainly holds licence data, but is aiming to expand over time to include carbon reporting data and carbon capture and storage records. The OGA also provides several tools for [browsing their datasets](#) and host an [open GIS data portal through Esri](#).

5.11.2. ADMIRALTY DATASET

[The ADMIRALTY Marine Data Portal](#) provides access to marine data sets held by the UK Hydrographic Office within the UK Exclusive Economic Zone (EEZ). This dataset/portal was created as a replacement for the original INSPIRE dataset, and as such follows the INSPIRE interoperability framework. It covers offshore infrastructure & energy assets, seabed compositions, shipping routes and a range of other marine data.

5.11.3. BRITISH OCEANOGRAPHIC DATA CENTRE (BODC)

The BODC is an oceanographic research focussed dataset, hosted by the National Oceanography Centre (NOC). The NOC as an entity holds an incredible amount of data, particularly concerning seawater composition, wave profiles, current flow, and specimens from over 100 years of marine research.

5.11.4. AEON PUBLIC DATA REPOSITORY (PDR)

The [Aeon PDR](#) is a free to access data repository hosted by Aeon Geoscience Systems. The PDR incorporates a vast database of spatial and subsurface data collated from the public domain and private databases. It aims to become a central repository for O&G/Offshore focussed data. Currently, it includes: regional geological maps, well logs, prospect / undeveloped discovery information and source rock studies at well & regional scales.

5.11.5. MARINE DATA EXCHANGE

[The Marine Data Exchange](#) is hosted by the Crown Estate. First developed in 2013, it aims to become a repository for offshore data that has been collected and is no longer commercially valuable. Their datasets are predominantly pre and post windfarm construction windspeeds, with some wave and tidal data.

5.12.OPEN DATA COLLECTIONS

Other than governmental data, there are a number of open datasets help by various GIS suppliers. These may also include some paid data, but by and large aim to share predominately academic data with the wider world.

5.12.1. ESRI

ESRI is one of the main players in the commercial GIS marketplace and host a large open data platform to go alongside their products. The [ArcGIS Hub](#) holds nearly 6000 datasets for all sorts of aspects of the UK, submitted by research organisations, the Oil and Gas authority, the UN, and many others. Energy specific datasets tend to be EV charger locations and usage and emissions reports.

5.12.2. SOAR.EARTH

There are several platforms competing with ESRI's hub, an example is [Soar.Earth](#), which aims to be a repository for projects to store GIS data, as well as providing a marketplace for providers to sell data to end users. A secondary aim for the platform is to create a service where drone operators can offer their services to create high resolution earth imagery on demand.

5.12.3. NATURAL ENVIRONMENTAL RESEARCH COUNCIL (NERC) DATA CATALOGUE

The [NERC](#) data catalogue is a portal that hosts a broad range of research related datasets, spanning earth, marine, polar, and atmospheric science, many of which are UK centric. Datasets span from historic (1800s) to the present day. Some specific topics that may be of use are changing land use and emissions data.

5.13. GIS DATASET LISTINGS (MIXED LICENSES)

Several places host lists of GIS data sources, one is maintained on [Wikipedia](#), and another by a [freelance GIS developer](#). Obviously, these may not be up to date, comprehensive, or accurate, but are good for getting a feel of what data is available. Much of it may not be directly applicable to energy projects, but often can be used to aid understanding on an area and add context to a problem. [Coastal Wiki](#) maintains a list of marine data portals, most of these are aimed towards ocean science, but does contain some links to subsurface data as well.

5.14. REGEN LOCAL ENERGY DATA INNOVATION REPORT

Regen, UK Research and Innovation, and the Energy Systems Catapult collaborated on a [research piece](#) into datasets that could be used to support energy innovation in local areas. The output is a tool that allows a user to explore data 'themes', and see how they interact and fit together within larger categories, as well as offering some insight into existing solutions in each category.

6. COMMERCIAL DATASETS

There are numerous other commercial GIS, and GIS related data suppliers. Obviously, not all will be able to be covered in this report, it aims to give a flavour of the types of data out there. Broadly, these can be split into two areas:

- High resolution satellite and mapping data.
- Computed/Pre-processed, proprietary datasets.

Within these areas, high resolution photographic data is the most competitive marketplace, with old players like Landsat competing against new market entrants. Both drone-based imagery, as well as 'traditional' satellites can offer resolutions of 5cm per pixel. Such high-resolution imagery could be very powerful in machine vision related applications and is currently being used to detect vegetation at risk of damaging powerlines. This kind of data is slowly being taken up and used more widely across various asset management workflows. Alongside photographic data, several companies offer imaging across a number of different spectrums.

Hyperspectral satellite imagery is not a new concept, but it has often been very low resolution, suitable for studying changes in river temperatures or mapping changing ice sheet thickness. As launching small satellites has become more accessible and sensors have improved, much higher resolution data is becoming available. An example is data detailed enough to produce building level thermal images. This kind of data beginning to appear on the market, which could be valuable in Local Area Energy Planning (LAEP) contexts to identify potential heat sources for district heat networks, to map out household insulation levels, or detect faults in power systems.

Pre-processed and computed data is typically centred around consumer data, and often uses UK census data as a baseline, with other, independent research mixed in. Many of these datasets are aimed at marketers, however, the best among them are often very accurate, finely grained, and clean datasets that can be used for any analysis the user sees fit. An example is [Experian Mosaic](#), which offers: "pre-packaged demographic, market-behaviour, and supply and demand models along with outlet locations at postal sector and output area levels." Such data could be used to derive insights into how innovative technologies are taken up in an area, and appetites for consumer spending, an important consideration in growing the uptake of clean energy technologies. Details of their datasets are available [here](#). Alongside Mosaic, there are services like [YouGov](#) which offer consumer datasets to the public and businesses freely and openly, albeit at a lower resolution than services like Mosaic, and Ipsos Mori's research.

APPENDIX – GIS DATA FORMATS

GIS data comes in a number of formats, some of which are semi-proprietary. This section will introduce the most common ones that apply to datasets in this document. For more information, [Esri's Github](#) has a good overview of other filetypes.

A1.1 Vector data and Raster data

GIS data can broadly be separated into two categories, raster and vector.

Vector data is the basis of most GIS data, as well as being used in computer aided design and digital illustration. Vector data and graphics uses geometric and mathematical constructs to represent 2D and 3D information. In practice, this allows a user to 'zoom' into data infinitely without losing detail. This also enables easier editing and reshaping of data than raster files. Vector data is computationally 'heavier' than raster data as all lines and shapes need to be calculated by the viewer. Due to this, vector formats are not suitable for storing continuous datasets, but are perfect for roads, powerlines, building shapes, and anything with discrete bounds.

Rasterised, or raster data, is a matrix of numeric values. A jpeg is a raster image, each 'pixel' (cell in the matrix) in the image is a value representing a colour, each one of which has a pre-determined size and shape, forming an image with a set size. This is why jpeg images become blurry when resized, each 'pixel' gets stretched or squashed. In a GIS raster format, each pixel can be tied to a location, allowing continuous datasets to be overlaid on a base GIS map, and matched up with points in a vector dataset. Gaps in raster data can easily be interpolated, and data can be transformed and combined with other raster datasets. Temperature, cloud cover, and distances from a point can be raster data.

A1.2 Comma Separated Value (CSV)

CSV files are one of the simplest forms of data used within GIS. They are a text file, which can contain almost any type of data, dependant on how the file is structured. In GIS, CSV's will most commonly be used to tie values to point data, but can be used to describe lines, polygons, statistical datasets, and all sorts of other specialised datasets. As a CSV can contain vector or raster data, it is important to carefully check the metadata associated with a CSV dataset and ensure any CSV data generated is well described.

A1.3 TIFF & GeoTIFF

A TIFF file is a raster image file format, commonly used for storing images, with associated data. This enables some complex features, such as character recognition and splitting of large images into multiple smaller files. GeoTIFF is a metadata standard that enables users to embed geospatial data into a TIFF image, allowing the image to be tied to a particular mapping system, projection, and co-ordinates.

A1.4 Shapefile

A Shapefile is format developed by Esri, and is one of the most common data formats used in GIS. The standard is open enough to be supported by most GIS software packages. It is a vector file format, that can contain points, lines, and polygons, and attributes about each feature. A shapefile is a collection of three separate files: *.shp* – the point data, *.shx*, an index of all the points to tell the software what to expect and how to order it, and a *.dbf* which contains all the attributes assigned to each point.

A limitation of the Shapefile format is that it does not support topology. Topology in GIS is the idea of connecting points together, which helps ensure there are no gaps between sections of line or polygon. From a data use perspective, this should not make much difference, however, it does mean datasets should be checked for any gaps if they may influence the analysis taking place.

A1.5 Geography Markup Language (GML)

A GML file is another standard for storing predominantly vector data, but with the option of extending it with a wide range of other data, including integrating raster data (termed coverage in the GML specification) into a single file. It is based upon XML, a widely used standard for encoding human and machine-readable files.

A1.6 GeoJSON

GeoJSON is a format similar to GML, but based on [JSON](#) rather than XML. It is particularly suited to simple features and small file sizes. This lends itself to defining simpler, abstract geometries and areas; GeoJSON is currently used within Apple's Mapkit, which requires a GeoJSON formatted file of the coverage a navigation app has. JSON as a whole is well suited to streaming data via an API and is therefore a good choice for data which needs to be accessed on demand. A recent extension to GeoJSON, TopoJSON can define geometries topographically, enabling more complex areas to be mapped and maintained without converting to other formats.

A1.7 Geopackage

A Geopackage is an open file format designed to support both raster and vector data, whilst being computationally efficient. The standard is extensible, allowing for innovative use cases, and is well suited for use in mobile applications due to small file sizes (compared to Shapefiles). Due to being a modern format, and still actively maintained and updated, they are easily interoperable with Python and R, as well as GIS packages.

A1.8 Network Common Data Form (NetCDF)

[NetCDF](#) (Network Common Data Form) is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. It is also a community standard for sharing scientific data. The Unidata Program Center supports and maintains NetCDF programming interfaces for C, C++, Java, and Fortran. Programming interfaces are also available for Python, IDL, MATLAB, R, Ruby, and Perl. NetCDF can be converted to rasterised GIS data, as well as used directly in machine learning models. This is a format commonly used for weather forecast data, which is a key driver of energy use.

A1.9 GRidded Binary (GRIB) Files

GRIB is a format defined by the World Meteorological Organization (WMO), a specialized agency of the United Nations. It is a simpler format than NetCDF, but has fewer dedicated libraries, creating a barrier to certain types of analysis. In some cases GRIB can be much more file size efficient, especially for very low-resolution data. GRIB1 is the original format and requires a separate parameter table to unpack the data. GRIB2 improves upon the standard with compression and the inclusion of a metadata/parameter table that you need to unpack the data in each file.

LICENCE/DISCLAIMER**Energy Systems Catapult (ESC) Limited Licence for ENERGY RELATED GIS DATA LANDSCAPE**

ESC is making this report available under the following conditions. This is intended to make the Information contained in this report available on a similar basis as under the Open Government Licence, but it is not Crown Copyright: it is owned by ESC. Under such licence, ESC is able to make the Information available under the terms of this licence. You are encouraged to Use and re-Use the Information that is available under this ESC licence freely and flexibly, with only a few conditions.

Using information under this ESC licence

Use by You of the Information indicates your acceptance of the terms and conditions below. ESC grants You a licence to Use the Information subject to the conditions below.

You are free to:

- copy, publish, distribute and transmit the Information;
- adapt the Information;
- exploit the Information commercially and non-commercially, for example, by combining it with other information, or by including it in your own product or application.

You must, where You do any of the above:

- acknowledge the source of the Information by including the following acknowledgement:
- "Information taken from ENERGY RELATED GIS DATA LANDSCAPE by Energy Systems Catapult";
- provide a copy of or a link to this licence;
- state that the Information contains copyright information licensed under this ESC Licence.
- acquire and maintain all necessary licences from any third party needed to Use the Information.

These are important conditions of this licence and if You fail to comply with them the rights granted to You under this licence, or any similar licence granted by ESC, will end automatically.

Exemptions

This licence only covers the Information and does not cover:

- personal data in the Information;
- trademarks of ESC; and
- any other intellectual property rights, including patents, trademarks, and design rights.

Non-endorsement

This licence does not grant You any right to Use the Information in a way that suggests any official status or that ESC endorses You or your Use of the Information.

Non-warranty and liability

The Information is made available for Use without charge. In downloading the Information, You accept the basis on which ESC makes it available. The Information is licensed 'as is' and ESC excludes all representations, warranties, obligations and liabilities in relation to the Information to the maximum extent permitted by law.

ESC is not liable for any errors or omissions in the Information and shall not be liable for any loss, injury or damage of any kind caused by its Use. This exclusion of liability includes, but is not limited to, any direct, indirect, special, incidental, consequential, punitive, or exemplary damages in each case such as loss of revenue, data, anticipated profits, and lost business. ESC does not guarantee the continued supply of the Information.

Governing law

This licence and any dispute or claim arising out of or in connection with it (including any noncontractual claims or disputes) shall be governed by and construed in accordance with the laws of England and Wales and the parties irrevocably submit to the non-exclusive jurisdiction of the English courts.

Definitions

In this licence, the terms below have the following meanings: 'Information' means information protected by copyright or by database right (for example, literary and artistic works, content, data and source code) offered for Use under the terms of this licence. 'ESC' means Energy Systems Catapult Limited, a company incorporated and registered in England and Wales with company number 8705784 whose registered office is at Cannon House, 7th Floor, The Priory Queensway, Birmingham, B4 6BS. 'Use' means doing any act which is restricted by copyright or database right, whether in the original medium or in any other medium, and includes without limitation distributing, copying, adapting, modifying as may be technically necessary to use it in a different mode or format. 'You' means the natural or legal person, or body of persons corporate or incorporate, acquiring rights under this licence.

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**