April 2023

Mobilizing private sector data for climate action

Prepared by Estelle Cantillon, Bertrand Collignon, Grégoire Denis and Guillaume Thys

Contact

Palina Shauchuk Palina.shauchuk@ulb.be

Disclaimer: This report was commissioned by the Data for Change – PARIS21 Foundation, with funding from the UK's Foreign, Commonwealth and Development Office, to provide a review of the state of the art of mobilization and use of private sector data for policy-making on climate change. It aims to foster a shared understanding of the opportunities, challenges and solutions surrounding their use, among the stakeholders of the data ecosystem for climate action (companies, policymakers, civil servants, civil society experts and academics). This final version reflects the discussions that took place at the report launch event in Geneva on March 24, 2023. The views expressed in the report are those of the authors and do not commit the PARIS21 Foundation, nor its funders.

Acknowledgments: This report has benefitted from the insights and advice of the following experts and practitioners: Sébastien Albert (Master Card), Stefan Dab (BCG), François Faelli (Bain & Co), Nastassia Lesczynska (ADE Evidence for Better Policy), Emmanuel Letouzé (DataPopAlliance), Frederic Pivetta (Dalberg Data Insights), Vikki Tam (Bain & Co), and Nicolas van Zeebroeck (ULB).

Citation: Cantillon, E., B. Collignon, G. Denis and G. Thys (2023), *Mobilizing private sector data for climate action*, Solvay Public Policy House, Report prepared for the PARIS21 Foundation, with support from the United Kingdom Foreign Commonwealth and Development Office

Table of Contents

Tables

Figures

1. Introduction

Addressing climate change requires effective and decisive action by all sectors of society. Data play a central role in supporting decision-making and fostering and coordinating action. The good news is that data have never been as plentiful as today. Likewise, thanks to advances in technology and software, our ability to analyse data has also never been as high as today.

A large part of data relevant for climate action is produced by the private sector. It is, therefore, important to **understand** to what extent these data are already or could be mobilized to support climate action and **identify what can be done** to further leverage these data for policy.

This report provides a starting point for this conversation. It provides a cartography of the main sources of private sector data and discusses their current and potential uses to support policy and public action in the context of climate change, with an eye on barriers, enablers, and emerging solutions.

For the purpose of this report, private sector data is defined as primary or secondary (repackaged) data produced by a non-governmental source. Private sector data are therefore characterized by their ownership, rather than their intrinsic characteristics. They include data produced by private (for or not-for-profit) organisations and data produced by citizens.

It is important to note at the outset that the definition of data on the basis of their ownership is not entirely without problems because ownership is context-dependent, and possibly changing, and, at the same time, not the primary criterion for use (the data attributes are). One example is satellite data, which are produced both by governmental agencies (NASA, ESA) and private for-profit entities. Another example are secondary datasets put together by private companies on the basis of public or, at least, open access primary data. The data become private, even though its origins are entirely public. Ownership has the merit of putting the question of access on the agenda. It does however run the risk of neglecting the more fundamental question: what kind of data do we need for the use case at hand ?

This caveat aside, we have identified three distinct features of private sector data that make them valuable complements or substitutes to existing public sector data. First, for some applications, private sector data offer a cheap and accessible alternative to standard public data (administrative data, census data and surveys). Second, some private sector datasets cover aspects and dimensions not covered by public sector data. Third, some private sector datasets have a level of granularity unmatched by public sector datasets.

Why private sector data ?

1. Cheap and accessible
2. Complementary to existing public sector data
3. Superior granularity

Does this mean that governments and the public sector are routinely using private sector data in the context of climate action? Of course not. Private sector data come in very different guises, and some are more useful than others, depending on the application. **Identifying those use cases where private sector data are indeed bringing real added value is a first challenge**. In the report, we identify the main private sector data types and discuss four areas where, in the context of climate action and especially in the Global South, private sector data can indeed bring value: impact assessments, policy targeting, citizen and community empowerment through data, and market development support. By combining data types with potential use cases, developing some in greater details through mini-case studies, we identify a typology of use cases and data types where private sector data are likely to be particularly valuable for climate action in the Global South.

A second challenge is to identify the best ways in which the public sector can tap into these private sector data. In this area too, there is no one-size-fits-all. Solutions range from mandated disclosure to data sharing partnerships and arms' length commercial relationships. In recent years, companies have realized the business value of the data produced as side products of their core business and many have set up divisions dedicated to repackaging these data for resale to third parties. Other companies have entered the business of matching, harmonizing, and repackaging private and public sector data from different sources. This means that, for many practical uses, the most convenient private sector data will actually be available on a commercial basis. For the remainder, we identify NGOs as potential effective data brokers for scalable datasets and summarize the findings from the literature on B2G data sharing partnerships for country or application-specific data.

2. What private sector data for climate action?

The range of data useful for climate is broad. Fossil fuel use, land use and agricultural activities are the main drivers of GHG emissions and therefore any data measuring these activities are relevant. When we think about the consequences of climate change, geophysical, demographic, and economic data are all relevant to quantify, prevent and/or alleviate them. Climate action is also fostering the development of new sectors, driving food systems change and, more generally, redefining economic advantage based on reduced environmental footprints of activities. Data are at the core of these transformations.

After screening the scientific literature, use cases, and company reports and websites, we identified the main types of data from the private sector (summarized in Table 1). Due to the high dimensionality of the data, several classifications can be established. Here, we classified data in three main categories: **Environmental and geospatial** data, **social and demographic** data, and **economic** data. For each data group, we provide its (i) production technology, (ii) type, (iii) measure, (iv) unit of observation, (v) temporal resolution, (vi) motivation to collect the data, (vii) typical primary owner, and (viii) typical access conditions for external parties. The production technology, type, and measure provide information on the data generation process: how they are produced (e.g. by satellites, by social media platforms, by transactions…), their type (sensing: a quantified measurement; content: a qualitative information; exhaust: the trace of an activity left by a user), and what measures they provide (e.g. geospatial data, opinion, position…). Next, the unit of observation (individual level, spatial resolution…) and the temporal resolution (sampling frequency) are important attributes that determine the suitability of a particular data type for a specific application. Finally, the motivation for sampling, the typical data owner, and the type of access (open access: data freely available, commercial access: access granted against payment; restricted access: data available after signing an agreement, usually a non-disclosure agreement or data sharing agreement) all provide indications about the current business model supporting the production of these data.

2.1. Environmental and geospatial data

Environmental and geospatial data are measurements related to the physicochemical (abiotic) or living (biotic) conditions and components of the environment or to human activities that impact the environment. This category includes satellite data, aerial imagery, and source-level environmental impacts. **Satellite data** are images of the Earth collected by imaging satellites. They provide geospatial data with a spatial resolution that can be as high as 0.3 square meter. These data are produced by governments or businesses, the later providing different access plans to the data (see also our Case study in section 4.1). **Aerial imagery** are geospatial data produced by planes or drones flying at lower altitudes. They provide more detailed mappings of the region that they fly over (spatial resolution up to a few square centimeters). Finally, we also included in this category **source-level environmental impacts**. These data compile all the measurements that are performed by companies or third-party auditors to quantify their emissions of effluents impacting the environment, with measurement-dependent technologies and resolutions, but also the impacts and outcomes of these emissions (e.g., biodiversity loss, land degradation, water quality…). Usually, these data are produced for their own operations, compliance or reporting purposes. In the first case, these data remain private (restricted access). In the second, they are either publicly available or stored in administrative databases. Note that we included in this category only the data that are collected and produced by activities, employees or contractors commissioned by companies, in opposition to crowdsourced data that are listed under "social and demographic data".

2.2. Social and demographic data

Social and demographic data are all the data that provide information on the circumstances in which people live or their activities, behavior, or opinion. This category includes social media, news media, web searches, data from telecommunications and GPS, connected wearables, and crowdsourced data. **Social media** data are any type of data that are gathered through a social media platform or application. They usually include both exhaust data (the trail of data left by using social media: connection time, location, or duration) and content data (content of a post, message, opinion…). These data are produced as a side product of the user's activity and/or collected for commercial use by the platform. Access is typically restricted, but several platforms have developed collaboration plans that allow access for non-commercial uses through a data sharing agreement (see our case study in Section 4.3). **News media** relates to the content of news, mainly to the general public. Monitoring the content of the news can provide valuable information about the current "hot" topics that are discussed, including on climate change and climate action. **Web search** data compile exhaust and content data connected to the online activity of internet users. These data include information from search engines (e.g., Google trends), visited websites, clicked links, time and location of connection and so on. **Telecommunication and GPS data** are exhaust data produced by the use of mobile devices (mainly mobile phones). Usually, these include metadata associated with the users' activity: position and timing of the communications (at the individual or tower level), number of communications, duration of communications, or social networks. As mobile phones have high penetration rates, they provide a large coverage of the population with a high level of granularity. **Connected wearables** (connected or "smart" watches, wristbands, pedometers…) are relatively recent technologies. These devices are mainly used as activity and health parameter trackers: they monitor the user's daily activities (e.g., number of steps, number of stairs…), heartbeat, quality of sleep, or sport performance. While these data are usually collected to enhance the users' experience (e.g., the possibility to track their progress), these data also provide detailed and individual medical records to private companies. Finally, **crowdsourced data** (or **citizen-generated data**) are data produced by participatory methods with the help of a large group of people (usually citizens). The dataset is thus built progressively by volunteers that feed it with local information that they have collected on their own. Examples include reports on pollution levels, fish catch or households' expenses, pictures, live reports of traffic accidents, and so on. While crowdsourced data can equally provide information environmental, geospatial, social, demographic and economic aspects, we chose to classify them in this section because they share many characteristics with other private sector social and demographic data due to their mode of collection. Specifically, these data can be noisy and rely on the ability of the volunteers to follow the same methodology for data collection. They can nevertheless provide valuable measurements for verification purposes, and highlight the different spectrum of real conditions that citizens experienced.

2.3. Economic data

Economic data are data that involve a monetary transaction or that are related to a specific market. **Transaction data** are data produced by financial transactions (purchases using a bank or credit card, payment with a phone, money transfers, insurance claims) or other purchases (fidelity programs). They typically contain information about the date, time, amount, currency, location (point-of-sale), and the parties involved in the transaction. Their representativeness depends on the data owner and type of transaction. A company such as Master Card covers transactions globally, but only credit card transactions. Banks, on the other hand, have access to all the transactions of their clients but their coverage is limited to their market. Collected at the individual level, these data provide very rich information about consumers, and, for this reason, most data owners are already commercializing anonymized versions of these data for marketing or market forecast purposes. **Electricity injection and withdrawals data** are collected by electricity transmission or distribution operators at the grid entry and exit points and by electricity retailers at the consumption points (smart meters). Telemetric versions of these meters enable frequent (15 min intervals) measurements that can be useful, for example, to detect consumption anomalies or encourage consumers (through dynamic tariffs) to adapt their consumption. **Production and sales data** cover company, product, point-of-sale or plant-specific production and sale metrics. These data typically stay within the company. Only a much-aggregated version is submitted to public authorities for tax compliance or other reporting obligation. Companies also communicate about their performances (e.g., in CSR reports). These reports and other **voluntary information sharing** by companies aim to engage with stakeholders. They are open access but there is no harmonization across companies concerning the type of information they report or their format. There is also typically no third-party verification of these data. Finally, a number of for-profit and not-for-profit actors compile and combine different sources of data for resale. A major typical benefit of these datasets is their multi-firm coverage (an entire sector or economy), all in harmonized and comparable format. Likewise, private certification bodies hold registries containing information on certified companies. These are classified under **third-party registries and databases** in Table 1. Depending on the data owners and motives for collection, these data are either publicly or commercially available.

Table 1: Typology of private sector data

	Production technology	Type of data	Measure	Unit of observation	Temporal resolution	Motivation	Typical data owner	Access
Environmental and geospatial data	Satellite data	Sensing	Geospatial data	0.3 square m and above	Near real time or days	Scientific Commercial use	Business Government	Commercial or open access depending on data and data owner
	Aerial imagery	Sensing	Geospatial data	A few square centimeters or less	Sampling-dependent	Commercial use	Business	Commercial
	Source-level emissions and impacts (effluents, water use, land degradation, biodiversity loss…)	Sensing	Measurement-dependent quantification	Measurement-dependent, from individual (or item) to household or company	Typically yearly (e.g. for emissions) or case dependent (e.g., one-shot assessment)	Scientific, Environmental assessment Own operations Reporting Compliance	Business	Open access, Restricted
Social and demographic data	Social media	Content Exhaust	Position, Movement, Content	Message, post, call	Real time or less	Side product of own operations Commercial use	Business	Commercial access to a subset, Restricted
	News media	Content	Content	Region / Country	Daily	Main product	Business	Open access, Restricted
	Web searches	Content Exhaust	Content	Keywords, website	Daily	Own operations	Business	
	Telecommunication and GPS	Exhaust	Position, Movement	Individual level, Call, message, Closest antenna	Real time or less	Side product of own operations	Business	Commercial access to a subset, Restricted
	Connected Wearables	Sensing Content	Physiology (heartbeat…), Activity (sport, steps…), Constitution (weight, height…)	Individual level	Real time or less	User experience, Commercial purpose	Business	Restricted
	Crowdsourced data (aka citizen-generated data)	Content	Images, pollution, economic activity,…	Specific to application	Irregular	Solicited for a specific application	Business NGOs Governments	Restricted
Economic data	Transactions (banking, insurance claims, fidelity programs, …)	Exhaust	Position, Timestamp, Amount	Transaction	Transaction rate	Side product of operations	Business	Commercial access to a subset, Restricted
	Electricity injection and withdrawal	Sensing	Physical	Electricity consumption block or point (smart meters)	15 min if telemetric	Own operations	Grid operator Retail company	Restricted
	Production & sales	Sensing	Volume, Price, Turnover	Product, POS Company	Daily or monthly	Own operations	Business	Restricted
	Corporate voluntary reporting	Variable	Content	Company	Yearly	Stakeholder engagement	Business	Open access
	Third-party registries and databases	Variable	Variable	Variable	Variable	Economic interest (e.g., certification) Scientific value	Business NGOs Academia	Open access, Commercial

Notes: Authors' compilation based on sampling the scientific literature, use cases, reports from public and non-profit institutions, and companies' websites.

3. Policy uses of private sector data in the context of climate action

The previous section provided a cartography of the near universe of data types produced by the private sector. A **complementary perspective** on the potential of private sector data for climate action is to start with the policy uses that these data could contribute to.

Due to the wide diversity of profiles and situations among the countries forming the Global South, there are many areas for climate policy action. Indeed, while the Global South as a whole is the leading emitter of greenhouse gases in absolute terms, the emission profiles range from one extreme (China and India are among the largest emitters in the world) to the other (almost all countries and territories among the 50 lowest emitters belong to the Global South). This disparity implies that some countries play a crucial role in global mitigation efforts, while other countries would mostly benefit from adaptation efforts (e.g., low emitters strongly impacted by climate change). In fact, many countries in the Global South are on the frontline of the climate crisis: they are highly exposed to climate change, and they combine a range of vulnerabilities making climate change an existential threat to their population and economy (Figure 1).

Figure 1: Aspects and dimensions of climate vulnerability in the Global South (regional averages of selected vulnerability indicators)

Source: IPCC (2022a), Figure TS.7 Vulnerability. This figure shows regional averages for selected aspects of human vulnerability. This regional information reveals that, within all regions, challenges exist in terms of different aspects of vulnerability, however, in some regions these challenges are more severe and accumulate in multiple dimensions.

Data can inform and improve public action at different levels of the policy cycle, from policy formulation to policy implementation and policy evaluation. Governments can also leverage data to support businesses, citizens and communities to take action. In this case, governments act more as enablers for climate action rather than as a central actor.

This section discusses four concrete policy applications where the role for data, and singularly, the role for *private sector* data in the context of climate action in the Global South, is high. For each policy application, we provide examples of existing uses of private sector data and highlight the data attributes that make these data useful for this type of application. These can be summarized in three key attributes (not all equally relevant for each application): granularity, complementarity with other existing data sources, cost effectiveness.

3.1. Impact assessment

Impact assessment seeks to identify and quantify the causal impact of an event or a policy. It is the data-intensive policy exercise *par excellence*. Impacts are typically multi-dimensional and a full assessment will therefore require a mapping of the potential direct and indirect impacts along all these dimensions. Given their richness (high frequency, high granularity), private sector can usefully *complement* public sector and other open access data in this exercise. This is all the more true that the typically one-shot nature of impact assessment does not require data access to be permanent and stable over time, a limitation of some of the restricted access private sector data.[1]

The following two examples illustrate how private sector data can complement public or open access data for impact assessment. In the summer 2022, Pakistan experienced major floods that covered one third of its surface. Crisis Ready, a multi-stakeholder organisation bringing together academics, non-profits, tech companies, legal experts and government, combined satellite data, high resolution administrative and spatial maps (GADM) curated by a researcher at UC Davis and social media data (Facebook) to track population movements during and after the floods and assess the impact on population density across administrative units.[2] The main benefit of social media here (but this could have been telecom data too) is their superior spatial and time resolution, relative to whatever an administrative census could have achieved. Another example is the use of card transaction data to evaluate the effectiveness of a voucher program introduced by the South Korean government to support consumption during Covid (Kim et al., 2020). The main benefit of credit transaction data here is the ability to assess how the constraints put on the vouchers led households to substitute away from non-eligible outlets, therefore providing some sense of the indirect impact of the program.

As an example of a policy assessment solely relying on public and private, but open access, data, ongoing work by ADE and the Center for International Forestry Research is evaluating the impact of an EU program to promote sustainable forest management and reduce illegal logging on local communities around protected areas in Cameroon.[3] In the absence of comprehensive livelihood surveys designed to specifically target this population, the researchers combine satellite data with geo-located livelihood and development data from the World Bank DHS to measure outcomes for the population of interest.

These examples illustrate a common imperative in impact assessments: the need to combine data from different sources, at the unit of analysis of interest (geographical area or household group for example). We return to this challenge in section 6.

3.2. Targeting

Policies may fail to reach their goals or induce unwanted redistributive effects because (some of) the intended beneficiaries miss the information to act upon. For example, schemes to support home insulation tend to benefit higher income households even if their homes are already among the better insulated. Some simple innovation may fail to be picked up by farmers or small firms for

[1] There are exceptions, of course, and for some applications (e.g., looking at impact of slow moving variables such as physical capital), one might be interested to track outcomes over the longer run.

[2] https://www.crisisready.io/2022/09/catastrophic-floods-devastate-southern-pakistan-crisisready-responds-with-new-data-reports/ (accessed March 2, 2023)

[3] More information about the program here: https://www.eu-flegt-vpa-programme.com/en/

lack of information. Relatedly, lack of information about individual circumstances can lead to costly undifferentiated policies because the public sector cannot do otherwise.

Targeting refers to the practice of identifying consumer, citizen or company profiles most suited for the intended message or policy. It is common practice in the private sector for marketing campaigns. In the policy context, targeting has been shown to increase take-up rates of energy-efficient appliances (Toledo, 2016) and reduce diversion of public subsidies (Banerjee et al., 2018), among others.

Effective targeting requires highly granular data on the population of interest to understand its living conditions or preferences. For example, geospatial data (from satellites or aerial images) can assess energy access and use and identify heat losses. Transaction data can provide insights into defensive or adaptation expenditures taken by households during heat waves. Knowledge about the fine structure of social interactions in a community (of the kind available for social media) can help speed up diffusion of the desired behavior and increase its eventual reach (Banerjee et al., 2016 and Banerjee et al., 2021).

3.3. Sunshine regulation and citizen empowerment

Dissemination of data to the general public can be a powerful tool for disciplining companies and empowering citizens and communities to address climate challenges. Sunshine regulation refers to the publication by public authorities of easily digestible data about the environmental, social or economic performance of companies, with the motivation that disclosure of such information can lead consumers, communities and other stakeholders to take action and put pressure on the underperforming companies.

Sunshine regulation refers to the publication by public authorities of easily digestible data about the environmental, social or economic performance of companies.

Sunshine regulation can be particularly useful in weak enforcement contexts where regulators may be absent or understaffed, and corruption rampant. In these contexts, information provides local communities, consumers, and other stakeholders the tools to discipline polluters and enforce social norms. Several effective sunshine regulation schemes have been implemented in the Global South (World Bank, 2000). More recently, consumers and investors in the North have started to play a role too and researchers have shown that environmental liability and human rights violations in supply chains weigh on sales and stock prices (Navarra, 2022, and Koenig and Poncet, 2022).

In the context of climate action, sunshine regulation, as a way to put pressure on polluters, is likely to be most effective in the presence of tangible co-benefits (on top of long-term global climate benefits) for current citizens and consumers, as these will facilitate collective action. Tangible present-day co-benefits of climate action include health (reduction of air pollution from the combustion of fossil fuels, reduction in noise, access to clean water), improved agricultural or fish yields from reduced pollution, better work conditions from adaptation measures, and jobs.

The benefits of climate-related data disclosure to the wider public go beyond their disciplining power on polluters however. First, information disclosure helps communities and citizens take **preventive and/or protective measures** (see e.g. Jia et al, 2019, Ito and Zhang, 2020, Jha and Nauze, 2022, in the context of air pollution but applications also include flood risks, selection of weather resilient crops, and so on). Second, information can **empower citizens to take climate mitigation action**. For example, a large number of phone apps exist today to inform consumers about the carbon footprint of their purchases and activities, or about the best time (from a climate perspective) to use energy-intensive electric appliances at home. Apps can also tell consumers about the state of the electricity grid and nudge them to shift their consumption if the grid is experiencing congestion due to power failure or peak loads.

So far, when describing the role of data for sunshine regulation and citizen empowerment, we have not specified their source, public or private. In practice, data ownership will depend on both the source of data and the existence or not of a business case, for a private entity, to collect and package them. Examples include:

- crowdsourced data collected through apps, website or regular mail (logs of consumer complaints, pictures of faulty behaviour or damages, …), repackaged into scores, warnings or any other easily accessible info. These could be collected by a public entity (e.g. the various fix-my-street apps), by an NGO as part of its societal mission (e.g. OpenStreetMap), or by a private ESG data vendor company as a complement to other sources of information;
- compliance data or any other source of open access data about company performance (such as news media, voluntary reporting, etc), suitably repackaged for accessibility by a public, private for-profit (e.g., RepRisk and Altana AI) or not-for-profit entity;
- satellite and aerial imagery.

An important requirement for these data to be used for sunshine regulation or citizen empowerment is **accessibility** and **consistency over time** so that the data can indeed be used for action and voice.

3.4. Market development support

One of the well-known challenges for climate action is the lack of incentives: individual mitigation action is costly but the benefits are global and in the future. This provides a rationale for policies such as carbon taxes, subsidies for energy efficient investments, emissions markets and standards, that either seek to realign individual incentives with social incentives or constrain behaviour. However, firms, cities and other organisations, are also increasingly voluntarily committing themselves to ambitious climate targets.[4] To meet them, they typically use a combination of own emissions reductions and carbon offsetting, i.e. reductions of carbon emissions or increase in carbon removals elsewhere.

Carbon offsetting creates a demand for carbon reduction projects. The potential for carbon reduction projects in the Global South is considered to be large and is the reason for the development of a specific framework, REDD+, within the United Nations. Some observers even consider that these could offer significant development finance for these countries. This could take the form of direct investment by impact investors or carbon offsets sold on voluntary carbon markets.

However, to date, these markets remain underdeveloped due to lack of transparency and trust. Prices are low (typically below 10 USD/ton). One of the main challenges is the difficulty to ascertain the quality of these projects and, in particular, the additionality and permanence of the resulting emissions reductions. Other challenges include the existence of competing standards and markets, cost of monitoring and evaluation, and low market liquidity and transparency.

[4] See e.g. https://sciencebasedtargets.org/companies-taking-action.

High data quality and technology are central to the development of the market for carbon offsets. Current methods for monitoring and verification rely on high spatial resolution satellite, ground-based sensing and artificial intelligence, and blockchain-based technologies are being developed to track and verify credits.[5] To foster transparency and support development, the Voluntary Registry Offsets Database at UC Berkeley aggregates data from the four largest credit registries. Several initiatives to encourage market standardizations are under way.[6]

This is all good news, but they will eventually increase the demands on data collection, verification, monitoring and reporting for carbon reduction projects. Governments can support project developers in their countries by facilitating their access to high quality standards labels.[7]

4. Case studies

4.1. Satellite data

Satellite data provide images of the Earth surface using different types of sensors (optical, thermal or radar), weather data (temperature, precipitation, wind speed), climate data (Earth's temperature, sea levels, ice coverage), environmental data (pollution levels, land use), and geophysical data (ocean currents, topography).

Satellite data are characterised by their spatial, temporal, spectral and radiometric resolutions. Spatial resolution refers to the level of detail with which a satellite can detect an object (the Earth surface captured in one pixel, for example 30 square meters). Temporal resolution refers to the frequency at which the data are collected for a specific location (e.g., daily, weekly). Spectral resolution refers to the ability of the sensor to define fine wavelength intervals (called bands) in the electromagnetic spectrum. This parameter is related to the level of detail with which each different surface features can be distinguished (for example, can it distinguish between different vegetation types?). For satellites used for land surface imaging, the share of high spectral resolution satellites (i.e., those providing more than 100 spectral bands) is relatively small, but increasing as the technology improves and the demand for more detailed information on the Earth's surface increases. Finally, radiometric resolution refers to the sensitivity of the sensor to differentiate different levels of brightness.

Table 2 provides a description of the possible climate-related uses of satellite data according to their spatial and temporal resolution. Monitoring of carbon sequestration and identification of areas with high potential for renewables are important for mitigation policies. High temporal resolution images are particularly well-suited to support climate adaptation efforts.

[5] See Mitchell et al. (2017) and Sipthorpe et al. (2022).

[6] See e.g., the Integrity Council for the Voluntary Carbon Market.

[7] The case for data to support market development goes beyond carbon markets and includes other nature-based financial products such as biodiversity credits and payments for ecosystems services.

Table 2: Satellite data resolution requirements according to uses.

	Spatial resolution		
Temporal resolution	Coarse (30 m or more)	Medium (10-30m)	Fine (< 10 m)
Low (≥ seasonable)		Glacier Monitoring	
Medium (weekly to monthly)	Monitoring of Carbon Sequestration in Forests and Ecosystems Ocean Monitoring	Coastal Zone Monitoring Urban Planning and Management	Identification of areas with high potential for renewables
High (≤ daily)	Agriculture Management Fossil fuel emissions	Land Cover and Land Use Change	Disaster Response Flood Monitoring Water Resource Management

Notes: Authors' classification on the basis of Alifu et al. (2020), Avtar et al. (2019), DeVries et al. (2020), Gómez et al. (2016), Haas et al. (2015), Hoque et al. (2017), Huang et al. (2018), Li et al. (2020), Martin (2014), Melet et al. (2020), Shanmugapriya et al. (2019) and Yu et al. (2023).

There were more than 5,465 satellites orbiting Earth as of 1st May 2022, with two thirds (66%) of those dedicated to communications (i.e. not producing data).[8] Most of the satellites used for Earth observation are owned by governments, with the US, China, Russia and Europe leading the pack. The largest private operators of satellites for Earth observation purposes are Planet Labs Inc., Spire Global Inc., and MAXAR. Table 3 describes the type of satellite data that the main operators are producing and their access conditions. All government-operated satellites provide global coverage and most private satellite operators have the ability to provide data for nearly all regions of the globe, depending on the specific mission asked. Many satellites are partially owned by multiple actors, including both government and private companies.

Table 3: Characteristics of primary producers of satellite data

	NASA	ESA	Planet Labs	Spire Global	MAXAR
Access	Free	Free	Subscription	Subscription	Pay-per-use
Categories of satellite data					
Images of Earth Surface	V	V	V		V
Weather data	V	V		V	
Climate data	V	V			
Environmental data	V	V	V	V	V
Geophysical data	V	V		V	V
Spatial resolution range for optical imagery	15m and above	5m and above	0.5m and above	-	0.3m and above
Temporal resolution range	0.5-16 days	Near real-time and above	Near real-time and above	Near real-time and above	Near real-time and above

Source: Authors' compilation on the basis of the websites of the respective organisation. Spatial resolution for NASA and ESA refers to the finest resolution of the images available for free. Spire Global operates only radar-based sensors.

In practice, end users rarely use primary satellite data. Instead, they source data from satellite data "repackagers" who clean the data (e.g., correction for cloud cover) and combine them with data analytics and visualisation software for easier use. All commercial satellite primary data producers

[8] Source : https://www.ucsusa.org/resources/satellite-database (accessed Feb. 26, 2023).

in Table 3 provide such added-value services. Examples of commercial satellite data repackagers include The Climate Corporation, a subsidiary of BAYER, that provides agricultural management services, and WeGaw, that helps energy companies optimize their hydro production. These private data vendors tend to concentrate their offers in the fine spatial resolution segment with a private sector customer base (energy companies, agrifood businesses). The cost of commercial satellite data varies widely depending on the type and volume of data. According to Statista, prices ranged from 14 to 25 USD/square kilometer in 2022 for one observation[9]. A daily refresh would then amount to several thousands of USD per square kilometer on a yearly basis.

An important segment of satellite data repackagers are research institutes and NGOs that produce user-friendly satellite data – typically tailored to one type of use – which they combine with other sources of data where needed to support their organisational goals. These near real-time data are then published in open access for anyone to use or under license forbidding commercial exploitation. For example, Global Forest Watch (GFW), an NGO funded and operated by the World Resource Institute, uses satellite data from NASA, ESA and private satellite companies to track and monitor changes in forest cover and carbon stocks (see side box). Global Fishing Watch combines open access location broadcast data produced by large vessels (called AIS for Automatic Identification System), government-sourced vessel monitoring data and geospatial satellite data from ESA and private satellite operators to monitor fishing, track human trafficking and support ocean management. Climate Trace brings together NGOs, tech companies and universities to leverage satellite data and artificial intelligence to produce real-time data on point-source GHG emissions.

The potential use of satellite data is still largely unexplored. Satellite data can be matched with any other geolocalized datasets to enrich the interpretation of the data and explore interactions with human and other ecosystems.

A number of initiatives exist to support countries in the Global South to access and use satellite data. The Global Environment Facility (GEF) provides funding to support the use of satellite data for sustainable development and climate action in developing countries. ESA has a program called Earth Observation for Sustainable Development (EO4SD), which provides financial support to countries

PERU'S USE OF GLOBAL FOREST WATCH DATA

The forests in Peru play a crucial role in the country's economic, social, and ecological systems. In order to combat deforestation, the Peruvian government has turned to Global Forest Watch (GFW) data, using it in a variety of applications to improve forest management practices.

One notable example is the "Alianza para la Conservación de la Amazonía" project, launched in 2017, which aimed to promote sustainable forest management and protect biodiversity in the region. GFW data was used to identify areas at high risk of deforestation, which allowed the government to prioritize conservation efforts accordingly.

Likewise, the Ministry of Environment in Peru used the GLAD alert system (jointly created by GFW and the University of Maryland) to identify at a 30-meter resolution likely areas of forest loss, before creating its own national alert system in 2017.

The Peruvian government also used GFW data in its "Operación Mercurio" initiative to identify areas where deforestation was taking place due to illegal mining activities. By using GFW data, the government was able to locate and shut down these illegal operations.

Source: Authors' compilation of information on the websites of Global Forest Watch, the University of Maryland School of Public Policy and Alianza Empresarial por la Amazonia.

[9] Source : https://www.statista.com/statistics/1293877/commercial-satellite-imagery-cost-worldwide (accessed Feb. 26, 2023).

in the Global South to help them use satellite data for sustainable development. The World Bank has several programs that provide funding for the use of satellite data in developing countries, including the Global Agriculture and Food Security Program. Additionally, a number of organisations provide technical assistance and capacity building support to countries in the Global South (see e.g. the Group on Earth Observations (GEO), a global network of government institutions, research institutions, data providers, businesses, engineers and scientists, UNOSAT, and the United Nations Office for Outer Space Affairs (UNOOSA)'s "Access to Space for All" program). These are complemented by a large community of users who are providing guidance in various forums.

4.2. Supply chains arrangements

Today, much of the production of goods and services is spread across several countries and locations in what are called Global Value Chains (GVCs). GVCs enable companies to optimize production for costs, speed, reliability, access to critical inputs and so on. Supply chains arrangements refer to the mechanisms put in place by companies to organise their supply chains and the distribution therein of information relevant for production, logistics and reporting.

Increasing stakeholder pressure, new reporting and due diligence requirements, as well as their own corporate social responsibility commitments, are forcing companies relying on GVCs to improve traceability and transparency in their supply chains. Traceability refers to the ability to track a product or material from its source through all stages of production, processing and distribution. It is a necessary requirement for evaluating the social and environmental footprint of a product. Transparency refers to the degree to which information about a company's supply chain is available and accessible to stakeholders.

These developments have direct implications for Global South countries. These countries are typically positioned in the lower rungs of GVCs, as suppliers of raw materials, intermediate goods and low-skilled labour. These are also the stages where the highest environmental and social impacts are typically concentrated. This means that some countries may be at the risk of being excluded from GVCs if their companies are not able to provide the data or meet the new standards.[10]

Existing approaches to sustainable supply chains address all the following considerations, albeit sometimes differently according to the nature of information shared and the sector:[11] (1) supply chain mapping, (2) product traceability (or supplier traceability), (3) data harmonisation, interoperability and standards, (4) transparency, (5) monitoring and auditing, (6) protection of commercially or reputation sensitive information, and (7) complexity and costs. Data harmonisation and interoperability can be challenging in long supply chains covering multiple impact dimensions, and standards can offer a cost-effective alternative in these cases. Transparency can be a powerful tool to leverage consumers' social and environmental concerns.

A critical concern in all supply chain arrangements is to ensure traceability while preserving commercially and reputation sensitive private information. Several solutions have emerged in practice. One solution is to associate a certificate with the product associated with the valuable quality and have market participants transfer this certificate along with the product. This is the

[10] See e.g., the concerns expressed in Presidential Climate Commission (2022).

[11] See Kashmanian (2017) for an overview with numerous examples.

model adopted by the Better Cotton Initiative, a multi-stakeholder initiative which sets standards for sustainable cotton, provide capacity building for farmers and traces cotton (through certificates) across the entire supply, from the farm to the finished garment. Another model, though still at its infancy, is to use blockchain technology to share data securely and ensure full product traceability. One example is IBM Food Trust, a blockchain-enabled network of growers, processors, wholesalers, distributors, manufacturers, retailers in the food sector. A third model, used for cross-country regulatory financial data, is to have a central data clearing agent that centralizes the data of all members of the supply chain but only shares the relevant aggregates to individual members (Barcellan et al., 2017). The closest to this model in supply chains is the Ellen MacArthur's Circulytics model to assess a company's performance in terms of circularity.

It is important to note that these data sharing arrangements in supply chains imply that, for all practical purposes, these extremely granular data are not accessible to third parties, let alone to a public sector entity.[12] However, there are private actors that can provide data and solutions that are sufficiently close to the intended data and can be useful for policy purposes (benchmarking of domestic suppliers, identification of sectors at risk of being excluded from GVCs). These include companies such Altana AI, Ecovadis, RepRisk and Trucost which all offer AI-powered data and analysis of supply chain risks and sustainability. There are also initiatives by industry federations to support data-sharing on supply chains and promote sector-level standards.

4.3. Social media data

Social media data is any type of information that can be collected on a social media or social network. In general, the term refers to the data gathered on individuals by social platforms that can count up to billions of users. These data can take many forms including text data, location data, images, surveys, connection time, or interaction networks among users. The data can be collected for internal operations (e.g. for development), explicitly gaining information (e.g. polls, surveys), or for commercial exploitation (e.g. profiling for marketing). Today, this process is usually done through an online platform (website) or an application on a mobile phone. Social media data are characterized by a timestamp, a

INTERNATIONAL PUBLIC OPINION ON CLIMATE CHANGE

To gain support for their policies, policymakers must be aware of the public level of knowledge and concern about climate change. In 2022, Meta shared with Yale researchers the survey results of 108,946 Facebook users from 192 countries/territories worldwide (NB: Facebook is banned from China so it was not sampled).

The results highlighted a high level of self-reported literacy about climate change in the Global North (GN), while a higher proportion of people from the Global South (GS) reported that they have "never heard" of it. The study also showed that the perception of risk was shared by the world population, but people in the GS were more likely to think that climate change would harm them personally. Likewise, most GS respondents stated that climate change was personally important. The world population considers that climate change should be a priority for their government, but GS respondents were less likely to think that they should reduce their GHG emissions regardless of what other countries do. When asked about the energy sector, a large majority showed a support for renewable energy, and users from the GN reported a high willingness to reduce fossil fuels consumption. Finally, most users (GN and GS) think that action to reduce climate change will improve or have no effect on the economy and job market.

Such study can help to develop impactful campaigns that address specific concerns of and provide targeted information to the considered population.

Source: Leiserowitz et al. (2022)

[12] One exception is the Carbon Disclosure Project which offers a hybrid model where companies submitting their data (GHG emissions, waters, forest) can decide to have them shared only with CDP signatories or made public.

location, a user profile, and a content. The timestamp and the location give access to a precise time-space coordinate that allow to geolocalize the users and identify their mobility patterns. The user profile allows to build a comprehensive profile of the user by matching successive information.

Finally, the content of the data (e.g., survey, message, picture, or post) provides intelligible information that can be analyzed and classified by modern algorithms and researchers.

More than 50% of the world population now possesses at least one user account on a social media platform. Table 4 summarizes the characteristics of the most used platforms. The six companies that possess the 10 most popular platforms are either based in the USA or in China. Most of them have developed a collaborative program for data sharing with non-commercial partners (researchers, academics, non-profit organizations): an application to access a specific dataset (e.g. Tweets from April and May 2022) can be submitted to a platform. The application is then evaluated by the platform that will provide the requested data to the applicant after signing a Data Sharing Agreement. In addition, some platforms (e.g. Twitter) offer the possibility to use an API (Application Programming Interface) that allows researchers to access the platform and search for specific data themselves (e.g. search for tweets that contain #ClimateAction). These API can come with different access levels and prices.

Table 4: Characteristics of the largest social media platforms (by monthly active users)

Platform	Monthly active users	Company	Country	Collaborative program for data sharing
Facebook	2.91 billion	Meta Platforms Inc.	USA	Yes (Data Sharing Agreement)
YouTube	2.56 billion	Alphabet Inc.	USA	Yes (Data Sharing Agreement)
WhatsApp	2.00 billion	Meta Platforms Inc.	USA	Yes (Data Sharing Agreement)
Instagram	1.48 billion	Meta Platforms Inc.	USA	Yes (Data Sharing Agreement)
WeChat	1.26 billion	Tencent Holdings Ltd.	CHINA	Not to our knowledge[13]
TikTok*	1.00 billion	ByteDance	CHINA	Yes (Data Sharing Agreement)
Facebook Messenger	0.98 billion	Meta Platforms Inc.	USA	Yes (Data Sharing Agreement)
Douyin*	0.60 billion	ByteDance	CHINA	Not to our knowledge
QQ	0.57 billion	Tencent Holdings Ltd.	CHINA	Not to our knowledge
Sina Weibo	0.57 billion	Sina Corporation	CHINA	Not to our knowledge but open API[14]
Twitter	0.45 billion	Twitter Inc.	USA	Yes (Data Sharing Agreement)

Notes. Authors' compilation of data from various sources including quarterly reports of social platforms. *Douyin and TikTok are the same social media platform that is split between a Chinese market and a rest-of-the-world marked.

We identified two main uses of social media data: sentiment analysis and mobility patterns.

Sentiment analysis

Sentiment analysis is the quantification of the affective states of users regarding a chosen topic. It relies on natural language processing and text processing to analyze the information content of posts or messages sent by users. While the technique can be applied to any topics, several studies have analyzed social media sentiment related to climate change on several social media

[13] Only a developer API exists for WeChat.

[14] Littman et al. (2017) provide a description of the Sina Weibo API.

platforms.[15] Thanks to the timestamp and geolocation of each post, it is for example possible to map the public sentiment on the platform according to the nationality of the user or track the evolution of the public opinion over time (see side box above).

Mobility and activity patterns

Mobility tracking aims at studying the movement of population over time. This can include short commuting route used on a daily basis (useful for city planning) or migration pattern in the long-term (e.g., displaced population in response to war or natural disasters). Most studies on mobility tracking use mobile phone data that usually provide a good temporal and spatial granularity of people on the move. However, as social media data also come with a timestamp and geolocation, they can also be used to study mobility and activity patterns.[16] The UN Refugees Agency estimates that 20 million people are displaced each year as a consequence of climate change. Being able to follow the movement of these populations over the years and across borders is thus crucial to develop a coordinated international response.

Benefits

Social media data can provide a rich and diverse set of perspectives and profiles with a high granularity (down to the individual level). As such, it can be a fast and cost-effective tool to sample a large population compared to traditional sampling methods (surveys, interviews). In addition, the data can be collected through time (temporal resolution) and by region (geographical resolution), regardless of the country that they are from or moving to. In this respect, they can be more useful for long term pattern analysis than mobile phone data that are dependent on specific operators that are usually territory-based.

Limitations

The main limitation of social media data is that the sample of users may not be representative of the population of interest. Even if they provide access to thousands of individuals (potentially billions), the use of social media is not distributed homogeneously among the population: younger and wealthier people tend to be more represented on social media. Likewise, people from different countries might use different social media platforms, with some platforms being banned from some countries (e.g., Facebook is banned in China, which tends to develop its own social media platforms). Social network data can also be contaminated by content polluters (e.g., bots, trolls) that automatically propagate political or advertising messages on the network. While some filtering methods exist, this pollution can distort the content analysis on a platform.[17] For these reasons, it is recommended to complement the data collected on social media with other datasets.

---

[15] See e.g. Effrosynidis et al. (2022a, 2022b), Mucha (2018), Fownes et al. (2018), An et al. (2014), Gaytan Camarillo et al. (2021) Sanford et al. (2019), Yeo et al. (2017), Falkenberg et al. (2022) and Debnat et al. (2022) for Twitter data, Spisak et al. (2022) for Facebook data, Zeng et al. (2022) for LinkedIn and Sina Weibo data, Hirsbrunner (2021) for YouTube data. Tuitjer and Dirksmeier (2021) and Mavrodieva et al. (2019) use combine several platforms.

[16] See the work of CrisisReady, Fraser (2022) and Hatchett et al. (2021) with Facebook data, the work of Yang et al. (2019) and Ebrahimpour et al. (2020) with Sina Weibo data, and Hasan et al. (2013) and Manca et al. (2017) with Twitter data.Researchers can thus follow individual movements through time or map the population change.

[17] See Nasim et al. (2018) for an example.

5. The case for private sector data

Figure 2: Four access routes to private sector data

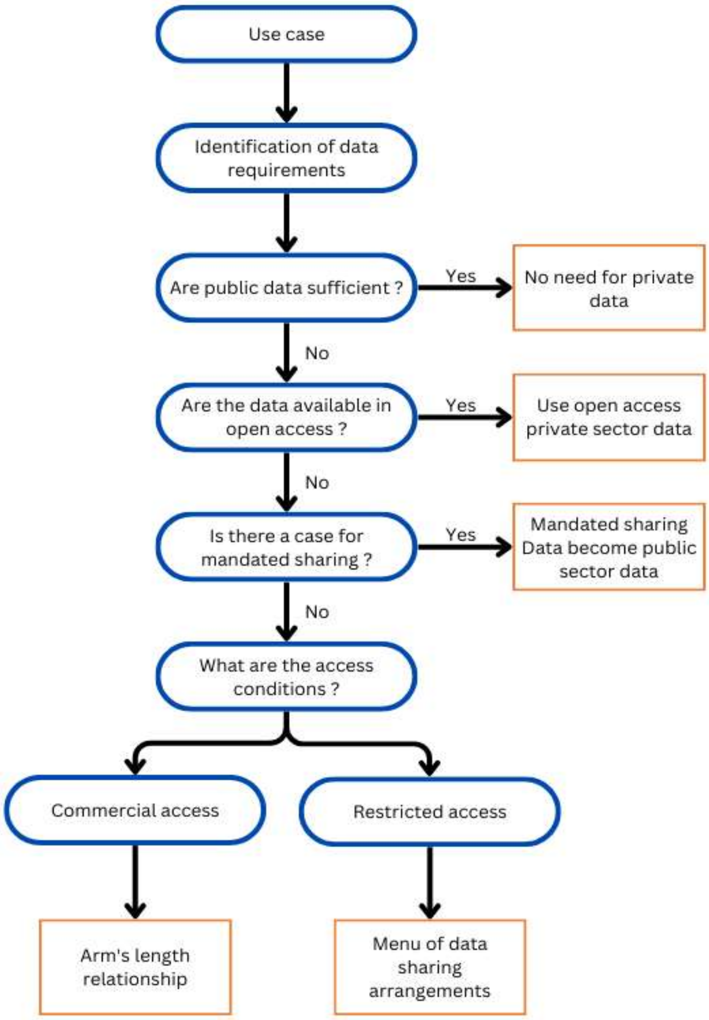
Our review of existing private sector data and their use for public policy has identified three major routes to the use of private sector data for policy: access to the data is open (possibly under a free license under the condition of non-commercial use), access to the data can be purchased, or access is restricted but the data owner is willing to share its data as part of a partnership or philanthropy. A fourth route, of course, is that of mandated sharing, where the public authority requests the private sector entity to share their data with the public authority. This is common practice in regulated sectors such as banking, health, and energy. In effect then, the data become administrative data and so no longer counts as private sector data even if they originate in the private sector. It is a route to consider if justified by the intended use of the data but is not our focus here.

Putting these two perspectives (use value and access) together, we propose to follow the decision tree shown in Figure 3 to determine which private sector data to use, if

From our mapping of private sector data, our discussion of data requirements for different policy applications, and our case studies, we can conclude that there is a case for the use of private sector data for policy and climate action.

However, it is also clear that the benefits from private sector data are case dependent. Compared to data collected by the public sector, private sector data can be more granular and cheaper, and they can provide complementary information (usually related to internal processes or private behaviours) not available in public databases.

Figure 3: Which private sector data, when ?



any, and what access route to select: (i) identify the data needs for the envisioned use case, (ii) identify those needs that cannot be covered by public databases (e.g., due to the scarcity or absence of public data, or their low resolution) and where, therefore, private sector data adds value, (iii) identify private data that could provide the missing information, and (iv) determine the framework to access these data (i.e., partnership, purchase, data-sharing agreement,…).

This decision process should give priority to the use of publicly available data when available, that can then be complemented by commercially available or restricted access private data once the benefits (added value, cost, ease of use, …) have been identified. Table 5 illustrates different potential use cases for private sector data in different policy areas related to climate action. For each policy area, we (i) provide one or two examples of use cases, (ii) describe the role of the data, (iii) the data requirements to meet these objectives, (iv) the justification for private sector data, and (v) the potential contribution to adaptation to or mitigation of climate change.

Table 5: Potential for private sector data for climate action (illustrative, non-exhaustive list)

Policy area	Illustrative use cases	Role of data	Data needs	Why use private sector?	Objectives (adaptation or mitigation)
Food Systems	Crop selection for culture and plantation	Identification of best crops, encourage adoption	Detailed land coverage, soil composition, and environmental conditions through satellite and aerial imagery	High resolution and cheap alternative to field measurements	Grow the most appropriate crops in response to changes in the environmental conditions (adaptation)
	Targeted fertilization	Define with precision the area where fertilizers should be applied	Detailed land coverage through satellite and aerial imagery	High resolution and cheap alternative to field measurements	Limit the amount of fertilizer that is applied and limit the impact of nutrient excess (mitigation)
Environmental monitoring and planet	Assessing the current state and resilience of natural systems	Revealing changes in soil quality, land coverage, water availability in natural ecosystems	Detailed land coverage, soil composition, and environmental conditions through satellite and aerial imagery and potentially	High resolution and cheap alternative to field-measurements	Quantify the impact of climate change and human activity on natural ecosystems to inform interventions in endangered zones (mitigation and adaptation)
Sanitation	Access to clean water	Identifying the population in need of clean water and the local quality of water sources	Population density (mobile phone or satellite data), Water quality	High granularity and cheap alternative to field-measurements	Ensure that populations have access to a clean source of water as a response to the threat on global water security. (adaptation)
Energy systems	Energy infrastructure mapping	Automated identification	Satellite or ground-based sensors	Cheap alternative to census	Optimize the energy distribution networks to reduce energy loss (mitigation) and respond to changing demand (adaptation)
	Access to electricity	Identify the territorial coverage of electric	Satellite (night lights)	Cheap alternative to census	One step towards a decarbonized energy (mitigation)
Transport infrastructure	Efficient transport network	Mobility tracking – most used route	Mobile data	High granularity and cheap alternative to field-measurements	Optimize the transport system to traffic needs and reduce CO_2 emissions (mitigation)
City planning	Traffic management (e.g., public transport and traffic lights)	Mobility tracking – most used route, peak hours	Mobile data	High granularity and cheap alternative to field-measurements	Optimize the daily commuting of citizens and reduce GHG emissions (mitigation)
	Park & Ride	Mobility tracking and land coverage	Satellite data Mobile data	Cartography of the city to plan the location of P&R services to meet the needs of commuters based on their mobility	Optimize urban transport planning and favor the use of alternative transports in the city, reducing GHG emissions (mitigation)

Notes: Authors' compilation from various sources including the Synthesis report Data-Pop Alliance (2015), Fetter and Baker (2020) for energy systems, Citilogik for city planning, and the sustainability report of the Coca Cola Company for its survey on water access and water quality.

6. Challenges and enablers for mobilizing private sector data

This section reviews briefly some of the challenges to private sector data access, and suggests that NGOs have an important enabling role to play for open access data, and that regulatory developments in the North can provide the ground for building the economic case for also considering commercial access as a mainstream route.

6.1. Challenges

Data harmonisation, standardisation and interoperability

Data harmonisation, standardisation and interoperability are a challenge for all access routes. Most if not all policy applications leveraging private sector data will pool data from different sources. Ideally, these data sources should be sufficiently harmonized and data variables and formats sufficiently standardized to allow linking and sharing. Standards and best practices in this respect do not differ from those that apply for public data (see e.g. UK Government, 2022) or scientific data (see e.g., the FAIR data Principles endorsed by the EU, Wilkinson et al., 2016), but the unstructured nature of some of the private sector data can make these issues particularly challenging. Techniques to enable linking include the well-established fuzzy matching algorithms which rely on similarities across the different datasets and more advanced imputing techniques based on artificial intelligence.[18]

Governance for data sharing partnerships

There is much interest and numerous pilots (some of which have discussed in earlier sections) of B2G data partnerships for data sharing when access is restricted (see e.g. Verhulst et al., 2019, Letouzé and Oliver, 2019, and Bartlomucci and Bresolin, 2022 for further examples). At the end of the day, however, this model will only be sustainable and scalable beyond one-shot demonstration cases if it meets the needs of the two sides of the equation: companies need to be willing to share their data and the public sector need to be willing to use them.

Table 6 summarizes some of the legal & governance, organisational, operational, technical, economic, and cultural barriers that have been identified in the existing literature for these B2G partnerships to develop outside of experimental sandboxes. We discuss them briefly in the following paragraphs.

[18] For example, a company like Fraym transforms household survey data into census-like spatial data.

Table 6: Barriers to B2G data sharing

	Private sector side	Public sector side	Global South specificity
Legal and governance barriers	Data protection regulations Lack of governance structure for secure sharing	Lack of governance structure for secure sharing License incompatibilities between datasets	Regulations of their enforcement can be weaker
Organisational barriers	Lack of KPIs and risk measures, shortage of dedicated staff, data sharing/collection/selling not part of every business model	Shortage of dedicated staff, capability	Shortage of dedicated staff even greater
Operational and technical barriers	Lack of available trusted operational & technical systems (posing various potential risks)	Lack of data interoperability, incomplete documentation, proprietary formats.	-
Economic barriers	Lack of incentives, high ex-ante transaction costs and perceived ex-post risks, potential negative impact on business value	Monopolistic data providers leading to potential high data pricing	Private-sector data often offered at no cost by telcos and big tech corporations; sharper potential negative impact on business value
Cultural barriers	General public not aware of potential for public interest	Lack of culture on data sharing, trust, slower adoption of digital transformation	-

Source : Authors' compilation based on Abraham et al. (2019), Alemanno et al. (2020), Barcellan et al. (2017), Bartolomucci and Bresolin (2022), Micheli (2022), George et al. (2022), Granell et al. (2022), Microsoft (2022), OECD (2020) and Verhulst et al. (2019).

Legal and governance barriers. Data protection regulations such as RGPD can constrain what companies can share and represent a first legal hurdle to B2G data sharing. Even when the data can in principle be shared, private and public sector entities lack governance frameworks to ensure they are securely shared. Data protection regulations, or at least their enforcement, can be weaker in the Global South, and has facilitated the development of B2G data sharing pilots.

Organisational barriers. Shortage of dedicated trained staff is a major organisational barrier for both public and private entities interested in B2G data sharing but, additionally, data sharing is typically not a core business of private entities. This means that decisions to enter data sharing agreements are treated on an ad-hoc basis, and therefore exposed to the usual inertia and status quo biases.

Operational and technical barriers. Once private and public partners have agreed on sharing their data, doing so may not be possible technically (lack of data rooms or secure connections) or the data are in a form that cannot be used by the recipient (proprietary format, lack of proper documentation, poor structure, large volume).

Economic barriers. The main economic barrier on the public sector side is the cost of acquisition of private sector data but it must be noted that companies typically adjust their commercial offer to their client type, and telecommunication companies and other big tech companies have so far shown a willingness to share their data for free to countries in the Global South. The main economic barrier to B2G data sharing therefore is the long-term business case for companies. B2G data sharing involves high transaction costs for private sector companies, little to no short-term financial rewards and leakage risks (with potential consequences on corporate reputation and the core business if commercially sensitive information is leaked). These perceived risks seem to be exacerbated for data sharing partnerships involving many partners and for data recipients located in the Global South (George et al, 2022). On the other hand, data sharing can also contribute to the emergence of data standards, and companies are increasingly seeing data sharing as a way to accelerate convergence in that area to reduce regulatory uncertainty.

Cultural barriers. Several authors note that the general public is largely unaware about the public interest of using private sector data for policy. This makes it more difficult to build the case internally for companies to share their data. On the other hand, the public sector may lack a culture of data sharing and may worry about the motives of private sector companies for sharing their data. In a sample of 12 case studies of private data sharing at the local administration level, Micheli (2022) found that, for this reason, local administrations preferred collaborative projects with private sector companies, rather than pure data transfers.

These barriers have led to a number of recommendations and emerging best practices summarized e.g. in Verhulst et al. (2019), Alemanno et al. (2020) (see side box) and Verhulst (2021).

RECOMMENDATIONS OF THE HIGH-LEVEL EXPERT GROUP ON B2G DATA SHARING (2020)

Governance

- Create national governance structures
- Create a data steward functions in public sector
- Encourage the use of sandboxes for data-sharing collaborations
- Explore the creation of an EU regulatory framework to facilitate public-sector reuse of privately held data for the public interest

Transparency and ethics

- Ensure transparency about B2G data-sharing collaborations
- Increase public awareness about the societal benefits of B2G data sharing
- Encourage the general public to share their data for public-interest purposes
- Develop ethical guidelines on data use
- Invest in data literacy of policy-makers and public sector workers

Operational models and tools

- Promote adoption of standards to reduce transaction cost and ensure interoperability
- Support the development of technologies to implement B2G data sharing at scale

Curation and maintenance of existing open access datasets

A number of the use cases that we have encountered writing this report make use of datasets produced and curated by NGOs, research centers or individual researchers. This was the case of the GADM dataset used to assess the impact of the massive floods in Pakistan in 2022 but there are many others. For example, Exiobase, a multi-country environmental input-output database tracking emissions and resource use by industry, the Emergency Events Database (EM-DAT), which tracks disasters worldwide since 1988, or the Berkeley Carbon Trading Project, which aggregates data from the four largest voluntary carbon credit registries. Typically, these datasets are fully open access or accessible for free for non-commercial uses.[19]

These cases nevertheless raise the concern about the preservation and maintenance of these datasets, once the specific funding is over, the NGO disappears or the researcher behind the data retires. A number of emerging practices seek to address these concerns at least for datasets produced by researchers. First, funding agencies are increasingly requiring researchers and other funded entities to develop a data management plan that includes hosting and preservation once the

[19] The reference to Creative Commons license standards to describe the exact conditions for use, remix, redistribution and adaptation is common.

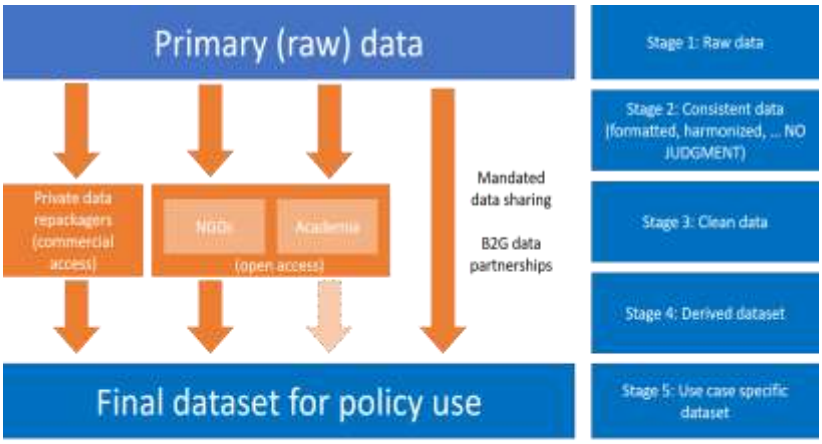
funding is over. Second, scientific journals are encouraging authors to register and host the data used for their research on third-party websites (e.g. Zenodo).[20] Registration provides many benefits including preservation and clarification of access conditions. Because they assign a DOI to the dataset, they also ensure proper credit assignment in future references, which can raise researchers' incentives to maintain the dataset. There is however no mechanism to fully guarantee that datasets that have clear value for policy are maintained. There is also no mechanism to ensure the preservation of datasets produced by NGOs.

6.2. Enablers

Anything that reduces the barriers identified in Table 6 will act as an enabler to greater mobilisation of private sector data by public authorities, not only in the context of B2G data sharing arrangements but also, when focusing on the public sector barriers, for any type of use of private sector data by public entities.

Our review of existing evidence and research has highlighted two potentially important enablers for private data mobilisation in the Global South: NGOs and climate and sustainability in the North.

Figure 4: Data stages and data intermediaries



NGOs as data intermediaries

There typically is a long way from raw data to usable data for policy. Raw (primary) data may be in a non-standard format, unstructured, partially redundant or, reversely, incomplete. The righthand side of **Error! Reference source not found.** describes the standard stages of data, from stage 1 (raw data) to stage 5 (the dataset used for the policy application). The Figure also shows the location, along these stages, of the ecosystems of the different access routes we have discussed. It highlights the key benefits of data intermediaries such as private data repackagers, NGOs and academia, who take care of the time-consuming tasks of data cleaning, harmonisation and preparation.

While commercial data access can be costly and datasets maintained by academics and research centres may not always be maintained beyond the life of the research project that motivated them, international NGOs can and do play an important go-between role. These NGOs have the human and technical capability to handle and analyse large complex datasets and package them for uses

[20] See e.g. https://social-science-data-editors.github.io/guidance/ or https://www.nature.com/sdata/policies/data-policies.

that support their missions. These NGOs are especially well positioned to mobilize global private sector datasets such as satellite data and social media. The global scale of these datasets means they can benefit from economies of scale both in handling the data and in providing solutions tailored to the local context. Looking back at Table 6, such go-between role can provide a solution to many of the existing barriers to B2G data sharing.

Climate and sustainability regulations in the North as enablers

Climate and sustainability regulations in the North increase the need for Global South countries to measure and track their natural resources and the climate impact of their companies that are integrated into global trade in order to be able to monetize their natural resources and continue to take part in global trade.

Supply chains disclosures. As part of their nations' commitment under the Paris Agreement, the European Union, the United States and other countries are increasingly encouraging companies located in their jurisdictions to disclose and decrease their emissions. Some of the new and upcoming legislations are also targeting these companies' entire supply chains. So, for example, France adopted a law in 2017 to impose a due diligence obligation for its companies to monitor, prevent and mitigate social and environmental risks in their supply chains. This includes carbon emissions. A similar directive proposal is currently under consideration at the EU level.[21] Likewise, the US Security and Exchange Commission recently published a proposal for climate-related disclosures for listed companies in the US.[22] The disclosure obligation would include scope 3 emissions, i.e. emissions produced in their supply chain. These developments mean that there will be an increasing need for measuring and tracking carbon emissions in supply chains, and an advantage for those suppliers in the Global South able to provide these data.

Climate-related finance. Additionally, and as discussed in section 3.4, the emerging market for climate impact projects, carbon offsets, biodiversity credits and related is data-intensive. Access of Global South countries to the high end of these markets can provide a good source of development finance but requires an ability to measure, trace and monitor natural capital consistently and at a high level of granularity.

Both of these developments increase the economic value of the relevant private sector data (satellite data, private databases, …) for companies and governments in the South, possibly making commercial subscription of some of these data worthwhile.

7. Proposed next steps

This report sought to provide an overview of existing practices and potential for mobilizing private-sector data for climate action, with a special focus on the Global South.

We have described the range of private-sector data and have identified the main attributes that make some of these data valuable for policy, namely (and depending on the use case) their low cost and accessibility, their complementarity to existing public sector data, and their superior granularity.

[21] Due diligence directive proposal, COM (2022) 71.

[22] SEC proposal for climate-related disclosures, RIN 3235-AM87, March 21, 2022

We have identified three existing routes to private sector data: commercial access, open access, and B2G data sharing agreements for restricted access data. There is today a rich offering of private sector data available commercially, and the offer is continuously expanding for ESG data. Most of these datasets have an international and multi-sectoral coverage and come with user-friendly visualisation and data analysis tools. There is also a rich offering of open access easily accessible geospatial data, provided by NGOs or academia, with policy support and development as their focus. These reduce the need to buy these data. What is left then for B2G sharing schemes are use cases that need granular socio-demographic data not available under the two other routes.

We think that all three routes are valuable and worth pursuing, but that the commercial access and the open access routes may have attracted less attention than warranted, given their low transaction costs and high scalability potential. In particular, we think that the low hanging fruit for capitalizing on private sector data might exactly lie on these two routes.

Consequently, our recommendations for the next steps are to:

1. Identify, by use case, best practices with at least the following criteria in mind: (1) reproducibility across locations and institutional environments, and (2) costs and scalability (are there economies of scale, learning-by-doing opportunities that will likely reduce these costs going forward?).
2. Identify climate policy relevant open access databases curated by NGOs and researchers and assess the value of linking and combining them with existing public databases.

8. References

Alemanno, A. et al. (2020). Towards a European strategy on business-to-government data sharing for the public interest. Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing European Commission.

Alifu, H., Vuillaume, J. F., Johnson, B. A., & Hirabayashi, Y. (2020). Machine-learning classification of debris-covered glaciers using a combination of Sentinel-1/-2 (SAR/optical), Landsat 8 (thermal) and digital elevation data. Geomorphology, 369, 107365.

An, X., Ganguly, A. R., Fang, Y., Scyphers, S. B., Hunter, A. M., & Dy, J. G. (2014). Tracking climate change opinions from twitter data. In Workshop on data science for social good (pp. 1-6).

Avtar, R., Sahu, N., Aggarwal, A. K., Chakraborty, S., Kharrazi, A., Yunus, A. P., ... & Kurniawan, T. A. (2019). Exploring renewable energy resources using remote sensing and GIS—A review. Resources, 8(3), 149.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, *341*(6144), 1236498.

Banerjee, A., Hanna, R., Kyle, J., Olken, B. A., & Sumarto, S. (2018). Tangible information and citizen empowerment: Identification cards and food subsidy programs in Indonesia. *Journal of Political Economy*, *126*(2), 451-491.

Banerjee, A., Chandrasekhar, A. G., Dalpath, S., Duflo, E., Floretta, J., Jackson, M. O., ... & Shrestha, M. (2021). *Selecting the most effective nudge: Evidence from a large-scale experiment on immunization* (No. w28726). National Bureau of Economic Research.

Barcellan, R., Nielsen, P. B., Calsamiglia, C., Camerer, C., Cantillon, E., et al. (2017). Developments in data for economic research, in: Blundell et al. (eds) Economics without Borders, Cambridge University Press

Bartolomucci, F., & Bresolin, G. (2022). Fostering Data Collaboratives' systematisation through models' definition and research priorities setting. In DG: The 23rd Annual International Conference on Digital Government Research (pp. 35-40).

Barwick, P. J., Li, S., Lin, L., & Zou, E. (2019). From fog to smog: The value of pollution information (No. w26541). National Bureau of Economic Research.

Data-Pop Alliance (2015). Big data for resilience: realising the benefits for developing countries. Synthesis report.

Debnath, R., Bardhan, R., Shah, D. U., Mohaddes, K., Ramage, M. H., Alvarez, R. M., & Sovacool, B. K. (2022). Social media enables people-centric climate action in the hard-to-decarbonise building sector. Scientific Reports, 12(1), 19017.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., ... & Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111(45), 15888-15893.

DeVries, B., Huang, C., Armston, J., Huang, W., Jones, J. W., & Lang, M. W. (2020). Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine. Remote Sensing of Environment, 240, 111664.

Ebrahimpour, Z., Wan, W., Velázquez García, J. L., Cervantes, O., & Hou, L. (2020). Analyzing social-geographic human mobility patterns using large-scale social media data. ISPRS International Journal of Geo-Information, 9(2), 125.

Effrosynidis, D., Sylaios, G., & Arampatzis, A. (2022a). Exploring climate change on Twitter using seven aspects: Stance, sentiment, aggressiveness, temperature, gender, topics, and disasters. *Plos one*, *17*(9), e0274213.

Effrosynidis, D., Karasakalidis, A. I., Sylaios, G., & Arampatzis, A. (2022b). The climate change Twitter dataset. Expert Systems with Applications, 204, 117541.

Falkenberg, M., Galeazzi, A., Torricelli, M., Di Marco, N., Larosa, F., Sas, M., ... & Baronchelli, A. (2022). Growing polarization around climate change on social media. Nature Climate Change, 1-8.

Fetter, R., & Baker, J. S. (2020). Harnessing Data Analytics to Accelerate Energy Access. NI PB, 20-03.

Fownes, J. R., Yu, C., & Margolin, D. B. (2018). Twitter and climate change. Sociology Compass, 12(6), e12587.

Fraser, T. (2022). Fleeing the unsustainable city: soft policy and the dual effect of social capital in hurricane evacuation. Sustainability Science, 17(5), 1995-2011.

Gaytan Camarillo, M., Ferguson, E., Ljevar, V., & Spence, A. (2021). Big changes start with small talk: Twitter and climate change in times of coronavirus pandemic. Frontiers in Psychology, 2308.

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. ISPRS Journal of photogrammetry and Remote Sensing, 116, 55-72.

Granell, C., Mooney, P., Jirka, S., Rieke, M., Ostermann, F., van Den Broecke, J., ... & Schade, S. (2022). Emerging approaches for data-driven innovation in Europe: Sandbox experiments on the governance of data and technology, JRC Technical Report (138 pp.)

Haas, J., Furberg, D., & Ban, Y. (2015). Satellite monitoring of urbanization and environmental impacts—A comparison of Stockholm and Shanghai. International Journal of Applied Earth Observation and Geoinformation, 38, 138-149.

Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In Proceedings of the 2nd ACM SIGKDD international workshop on urban computing (pp. 1-8).

Hatchett, B. J., Benmarhnia, T., Guirguis, K., VanderMolen, K., Gershunov, A., Kerwin, H., ... & Samburova, V. (2021). Mobility data to aid assessment of human responses to extreme environmental conditions. The Lancet Planetary Health, 5(10), e665-e667.

Hirsbrunner, S. D. (2021). Negotiating the data deluge on YouTube: practices of knowledge appropriation and articulated ambiguity around visual scenarios of sea-level rise futures. Frontiers in Communication, 6, 613167.

Hoque, M. A. A., Phinn, S., & Roelfsema, C. (2017). A systematic review of tropical cyclone disaster management research using remote sensing and spatial analysis. Ocean & Coastal Management, 146, 109-120.

Huang, C., Chen, Y., Zhang, S., & Wu, J. (2018). Detecting, extracting, and monitoring surface water from space using optical sensors: A review. Reviews of Geophysics, 56(2), 333-360.

IPCC (2022), Climate Change 2022 : Impacts, Adaptation and Vulnerability, Working Group 2 of the Sixth Assessment Report.

Ito, K., & Zhang, S. (2020). Willingness to pay for clean air: Evidence from air purifier markets in China. Journal of Political Economy, 128(5), 1627-1672.

Jha, A., & Nauze, A. L. (2022). US Embassy air-quality tweets led to global health benefits. Proceedings of the National Academy of Sciences, 119(44), e2201092119.

Koenig, P., & Poncet, S. (2022). The effects of the Rana Plaza collapse on the sourcing choices of French importers. Journal of International Economics, 137, 103576.

Koh, K., & Lyou, W. (2020). Spending impact of COVID-19 stimulus payments: Evidence from card transaction data in South Korea, Singapore Management University, working paper 9-2020

Leiserowitz A, J. Carman, N. Buttermore, L. Neyens, S. Rosenthal, J. Marlon, J.W. Schneider and K. Mulcahy (2022), International public opinion on climate change, New Have CT: Yale Program on Climate Change Communication and Data for Good at Meta

Letouzé, E., & Oliver, N. (2019). Sharing is caring: four key requirements for sustainable private data sharing and use for public good. DataPopAlliance, NYC.

Li, W., Niu, Z., Shang, R., Qin, Y., Wang, L., & Chen, H. (2020). High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. International Journal of Applied Earth Observation and Geoinformation, 92, 102163.

Littman, J., Kerchner, D., He, Y., Tan, Y., & Zeljak, C. (2017). Collecting social media data from the Sina Weibo API. Journal of East Asian Libraries, 2017(165), 12.

Manca, M., Boratto, L., Roman, V. M., i Gallissà, O. M., & Kaltenbrunner, A. (2017). Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study. Online Social Networks and Media, 1, 56-69.

Martin, S. (2014). An introduction to ocean remote sensing. Cambridge University Press.

Mavrodieva, A. V., Rachman, O. K., Harahap, V. B., & Shaw, R. (2019). Role of social media as a soft power tool in raising public awareness and engagement in addressing climate change. Climate, 7(10), 122.

Melet, A., Teatini, P., Le Cozannet, G., Jamet, C., Conversi, A., Benveniste, J., & Almar, R. (2020). Earth observations for monitoring marine coastal hazards and their drivers. Surveys in Geophysics, 41, 1489-1534.

Micheli, M. (2022). Public bodies’ access to private sector data: The perspectives of twelve European local administrations. First Monday, 27(2).

Microsoft (2022), Closing the climate data divide gap in the global south, https://blogs.microsoft.com/on-the-issues/2022/11/07/climate-data-divide-global-south/ (accessed Feb 27, 2023)

Mitchell, A. L., Rosenqvist, A., & Mora, B. (2017). Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. Carbon balance and management, 12(1), 1-22.

Mucha, N. (2018). Sentiment analysis of global warming using twitter data. Master thesis.

Nasim, M., Nguyen, A., Lothian, N., Cope, R., & Mitchell, L. (2018). Real-time detection of content polluters in partially observable Twitter networks. In Companion Proceedings of the The Web Conference 2018 (pp. 1331-1339).

Navarra, E. (2022). Stock Market Response to Firms' Misconduct, Available at SSRN: https://ssrn.com/abstract=3930980

OECD (2020), Mapping approaches to data and data flows, *Report for the G20 Digital Economy Task Force.*

Presidential Climate Commission (2023), Carbon Border Adjustment Mechanisms and Implications for South Africa, February, available online.

Sanford, M., Painter, J., Yasseri, T., & Lorimer, J. (2021). Controversy around climate change reports: a case study of Twitter responses to the 2019 IPCC report on land. Climatic change, 167(3-4), 59.

Shanmugapriya, P., Rathika, S., Ramesh, T., & Janaki, P. (2019). Applications of remote sensing in agriculture-A Review. Int. J. Curr. Microbiol. Appl. Sci, 8(01), 2270-2283.

Sipthorpe, A., Brink, S., Van Leeuwen, T., & Staffell, I. (2022). Blockchain solutions for carbon markets are nearing maturity. One Earth, 5(7), 779-791.

Spisak, B. R., State, B., van de Leemput, I., Scheffer, M., & Liu, Y. (2022). Large-scale decrease in the social salience of climate change during the COVID-19 pandemic. *Plos one*, *17*(1), e0256082.

Toledo, C. (2016). Do environmental messages work on the poor? Experimental evidence from brazilian favelas. Journal of the Association of Environmental and Resource Economists, 3(1), 37-83.

Tuitjer, L., & Dirksmeier, P. (2021). Social media and perceived climate change efficacy: A European comparison. Digital Geography and Society, 2, 100018.

UK government (2022), Data sharing governance framework guidelines

Verhulst, S. G., Young, A., Winowatan, M., & Zahuranec, A. J. (2019). Leveraging private data for public good. Govlab.

Verhulst, S. G. (2021). Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs. *Data & Policy*, *3*, e6.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1-9.

World Bank (2000), Greening industry : New roles for communities, markets and governments, Oxford University Press.

Yang, C., Xiao, M., Ding, X., Tian, W., Zhai, Y., Chen, J., ... & Ye, X. (2019). Exploring human mobility patterns using geo-tagged social media data at the group level. Journal of Spatial Science, 64(2), 221-238.

Yeo, S. K., Handlos, Z., Karambelas, A., Su, L. Y. F., Rose, K. M., Brossard, D., & Griffin, K. (2017). The influence of temperature on# ClimateChange and# GlobalWarming discourses on Twitter. Journal of Science Communication, 16(5), A01.

Yu, X., Millet, D. B., Henze, D. K., Turner, A. J., Delgado, A. L., Bloom, A. A., & Sheng, J. (2023). A high-resolution satellite-based map of global methane emissions reveals missing wetland, fossil fuel, and monsoon sources. Atmospheric Chemistry and Physics, 23(5), 3325-3346.

Zeng, L., Li, R. Y. M., Mao, Y., Chen, H., & Zeng, H. (2023). A comparative study on LinkedIn and Sina Weibo users' perceptions of the carbon-neutral city. Social Media, Artificial Intelligence and Carbon Neutrality, 16648714, 106.