

EU Company Data: State of the Union **2020**

How poor access to company data is
undermining the European Union

Executive summary	4
Why access to company data matters	6
The fundamental business dataset	6
The EU context: 27 member states, one Market	7
1. Non-EU specific issues	7
2. Societal and business issues specific to the EU	9
The EU Open Data and PSI Directive	15
The results: overview	17
What good looks like	17
What bad looks like	18
How does this compare with other countries?	19
A tale of two states: Sweden vs Denmark	20
Other notable issues relating to EU company data	20
Publishing frequency	20
Licences	20
Identifiers	21
What is a company (and other data quality and standardisation issues)?	23
GDPR and privacy	24
Foreign branches: when is a company not a...	27
Black Box data: why it's problematic, and how closed registers enable it	27
Bulk data vs API: which is better?	28
Bulk data use cases	29
API use cases	29
Keeping bulk data fresh	30
The real-time, computer-controlled future of company formation	30
Why charging for company register data is so problematic	31
Benefits open company data provides for EU countries	32
Direct benefits to EU member states	32
Indirect benefits to EU member states	34
Detailed results	36
Methodology	37
Discovery	37
Scoring	37
1. Unrestricted online search (no cost, no registration, search feature)	37
2. Openly licensed	38
3. Free machine-readable data	38

4. Data depth – directors	38
5. Data depth – annual accounts	39
6. Data depth – shareholdings	39
About OpenCorporates	39
Errors And Corrections	39

Executive summary

This report shows the current state of play in access to official company data in the EU, country by country, and also details the problems in the current environment.

- **Poor overall performance – exemplified by Spain (0), Austria (0) & Italy (10)**
 Overall the EU still performs very badly on access to official company register data, scoring just 40 out of 100, with two member states, most notably Spain and Austria, scoring 0 out of 100, and Italy scoring just 10/100.
- **Little improvement in the past 8 years**
 In the 8 years since we last reported on access to company data in the European Union there has been some improvement, but data releases have been limited, piecemeal and of variable quality.
- **No progress at all in the past 3 years**
 There has been no progress at all in the past 3 years, despite an increased awareness of the importance of data in society and business, and an understanding of the role that imbalances of access to it presents – both at an economic and societal level.
- **EU single market is undermined by poor access**
 The whole point of the EU single market is to reduce cross-border barriers to competition, trade and innovation. Yet access to the fundamental business dataset – companies and company ownership – is restricted not to those who most need to use it, but to those who can afford to pay the very considerable cost, undermining both the practice and the concept of the single market.
- **Illicit activity is free to proliferate whilst access is poor**
 Reducing access to this dataset enables money laundering, organised crime, corruption and other illicit activities.
- **Corporate accountability deficit widens – powered by EU Covid-19 financial aid**
 There is a significant corporate accountability deficit in the EU, and this looks set to widen over the next few years, powered by technology and, ironically, the trillions of Euros given to business as part of the Covid-19 pandemic response. This will undermine trust in democratic institutions and the social contract.
- **Efforts to restrict access persist – and are growing in some states**
 In several countries there appears to be a hard policy of restricting access to company register data to those with money, and in others there are moves to further restrict the use of the data, including the use by civil society for investigations.
- **Open Data Directive is opportunity for improvement, despite efforts to undermine it**
 There is some light on the horizon with the 2019 PSI and Open Data Directive, which is due to be implemented in the summer of 2021. This categorises company registers and company ownership information as high-value datasets that must be made available as open data. However, there is considerable pushback from a small number of member states, and it seems possible that the directive will be undermined by the addition of restrictive terms, delays and carve-outs.

Why access to company data matters

The fundamental business dataset

Why does access to company register data matter so much? Companies are artificial entities (as opposed to ‘natural persons’, i.e. people) given legal personality by the state for the wider benefit of society. These legal constructs (often called legal entities) can hold assets, owe money, enter into contracts, and break the law – yet the company itself, as opposed to the buildings it owns or the people it employs, has no physical form. You can’t touch it, or put it in handcuffs.

Because of this, companies are extraordinarily versatile, and can be used for a wide variety of purposes. These range from traditional uses such as large corporations and small business, to less savory ones – such as the facilitation of money laundering and organised crime.

Consequently, this dataset – the list of companies that have been incorporated in a jurisdiction and the key facts about them – is at the heart of all business, and, increasingly, our day-to-day lives.

On a business level, most organisations have hundreds of interactions with different companies every day – and not just traditional ones, such as with suppliers, clients, banks, lawyers and accountants. They also increasingly include so-called Software as a Service (SaaS) platforms such as Salesforce, Xero, Stripe, GitHub, Mailchimp or Slack – not to mention interactions via social networks and online advertising.

The situation is no less complex on a personal level. We interact with companies a hundred times a day – in shops, at work, at play, via each of the different apps on our phone, through the myriad websites we navigate, through social networks, through the ads we see, and even through smart speakers and home appliances.

That’s why it is so important that everyone – employees, suppliers, customers, governments, banks, competitors, journalists – can answer the question: *‘Does this company exist, and what are the core, statutory facts that define it?’* In our data-driven, app-centric connected world this means having the data where, how and when we need it.

This is what we mean by ‘core company data’: the minimal set of data about a company that is needed for free, open, fair and stable markets to exist, and for society to understand, influence, and, where necessary, regulate the behaviour of companies.

The EU context: 27 member states, one Market

At the heart of the European Union is the concept of a single market – in goods, in services, in labour and in capital. This means, among other things, that any company or person can do business in another member state as if they were in that state – the so-called Freedom of Establishment¹. This is intended to reduce barriers to competition, improve prosperity and encourage innovation.

It can also potentially enable a race to the bottom – in tax, regulation and transparency – and this, among other reasons, is why full and free access to official company data is so crucial in an EU context.

This report examines in detail the reasons why open company data is so important from an EU perspective, and the issues caused by the lack of it. For ease, we have split these into three areas:

1. Non-EU specific issues
2. Societal and business issues specific to the EU
3. Technical issues specific to the EU

1. Non-EU specific issues

Today, all business is to some degree global. Even if a business serves a specific local area with no website, it will compete with global companies, have global suppliers, or operate within laws and regulations shaped by the global, connected world in which we all live and work.

Yet information about companies is heavily siloed, non-standardised and rarely available as open data, even in a single market area such as the EU or the USA – despite it clearly being better for business and for citizens. Together these two issues – siloed non-standardised data and difficulty accessing it as open data – have had significant implications:

- **Proprietary data has proliferated**

When cross-border company data is needed (cross-border trade is after all the point of a single market), the gap has been filled by expensive and opaque proprietary vendors, whose business model depends on restricting access to all but the relatively small number of users that can afford their services.

This has increased costs and work, reduced utility and introduced frictions and inefficiencies that distract businesses, who are the primary users of company register data, from the real work they should be doing. It also locks out significant sections of society and introduces data quality issues (see *below*).

¹<https://www.europarl.europa.eu/factsheets/en/sheet/40/freedom-of-establishment-and-freedom-to-provide-services>

- **Not fit for purpose in a data-driven world**

The current lack of access to official company register information as data is highly problematic in our data-driven world, where innovation and insight comes from combining datasets, or looking at the records not one at a time but as whole datasets. Whether you are identifying innovation and entrepreneurship hotspots, improving risk metrics, investigating corruption or analysing gender inequality in board composition, today's world requires datasets, not documents.

- **Official public data is underused**

When official company register data is locked away in siloed databases, it is undermining the very purpose of the register – to be something that is consulted. Today, in our connected, online world, this means being made available when, how and where it's needed – whether this is via mobile apps, Software as a Service (SaaS) apps, procurement systems, or investigative dashboards. When the UK's company register, Companies House, opened its data via open data downloads and free APIs, it increased its usage by a huge amount – well more than ten times.

- **Poor data-quality feedback loops**

There are multiple data-quality issues in the current siloed system, both on the individual record level and systematic ones. These persist due to poor data-quality feedback loops and by the inability to combine datasets together. An example of the former are director records that are wrong or out of date, unnoticed by the directors concerned due to the lack of visibility of the data, or data on foreign branches that conflicts with home company records (see *below*).

- **Financial crime proliferates**

The current situation – siloed non-standardised, non-open data – provides a fertile ground for tax evasion, fraud, money laundering and organised crime, much of which happens using legal entities across multiple jurisdictions. The work of law enforcement, tax officers and regulators (as well as journalists and civil society organisations) is made significantly more difficult in the current environment as they are, in practice, restricted to data from their home country.

- **Government procurement is opaque and inefficient**

There have been significant improvements in the transparency of government procurement in the EU over the past 10 years. These improvements have been aided by EU and national initiatives, as well as by work by civil-society organisations such as the Open Contracting Partnership. However, as work being done by the TheyBuyForYou project² (funded by an EU Horizon 2020 grant) shows, the lack of open company data makes understanding this data considerably more difficult. Even identifying the recipients of government contracts is difficult, never mind answering relatively simple questions such as "Who are the biggest suppliers to a government?".

² <https://theybuyforyou.eu/> Note: OpenCorporates is one of the partners in this project

Nor are these difficulties limited to external stakeholders, but they are endemic within government too, raising costs and perpetuating inefficiencies.

- **Other stakeholders are shut out**

Business is not the only group that needs access to official company data. Everyone from journalists, unions, employees, or just ordinary members of the public need access. All users of Facebook or Google, for example, should be able to see the corporate structure of these huge corporations. Similarly, anyone buying a drink from a coffee shop should be able to see the shop’s beneficial owners (is it really a little independent, or controlled by a multinational, or even used as a front for money-laundering?). These other stakeholders have even been recognised by the Business Roundtable (an organisation made up of the Chief Executives of the largest corporations in the US), which [last year made clear](#) there are multiple stakeholders of companies – not just shareholders, but also customers, employees, suppliers and communities. The opacity that the current siloed, closed access brings effectively shuts them out.

2. Societal and business issues specific to the EU

As noted above, the single market is at the very heart of the EU, and anything that undermines this or the policies of the Union, is highly problematic.

Some of these barriers are for historic reasons as the registers in Europe have many varied histories – even company incorporation varies significantly across the member states. Others appear more deliberate, such as registers restricting access to their data for no good public interest reason, or protecting the relatively tiny income of the register at the enormous expense of the wider market.

We have identified 15 specific areas where the lack of company register data as open data frustrates, impedes or otherwise has a negative impact on EU policy (these are in addition to the 7 general categories listed above).

The cost of these issues runs to many billions of Euros for the member states and their citizens, and more than that stifle the innovations of its companies and undermine confidence in the single market, as well as national and EU institutions.

1. **Supply chain integrity and fairness**

There have been many scandals uncovering bad practices in supply chains that it is difficult to know where to start – food scandals, human rights abuses, corruption and bribery, environmental damage, counterfeiting, and unfair practices, especially larger companies. Underpinning these issues are two fundamental problems. One is a lack of clarity companies have on their own supply chains (these are, by definition, a chain of contractual relationships between legal entities), including a lack of visibility on exactly

what the companies are, and who is behind them (and hence their past record). The second is the opacity of such chains, and the lack of leverage that consumers have to effect change³ on them. For both of these, open company data, particularly ownership and directors, and corporate structures is crucial.

2. **Corporate accountability**

There are no specific directives or initiatives that we know of by the EU on corporate accountability. However, there are numerous statements made by the European Commission and EU member states on this subject. For example, the European Commission statement on the Single Market Act said⁴, “It is of paramount importance that European businesses demonstrate the utmost responsibility not only towards their employees and shareholders but also towards society at large.”

But accountability is not possible without access to information, and, in this world where, increasingly, power comes from the ability to analyze, reuse and combine large datasets with other large datasets, that means access to the information as data in a machine-readable form under an open licence that allows reuse.

3. **Reducing unnecessary burdens on business through regtech and fintech**⁵

The fintech and regtech markets are growing apace, and have the potential to significantly reduce the burden on businesses, while improving compliance and standards. However, like most B2B markets, core company data is a key foundational dataset in fintech and regtech, and by restricting access not only is this growing sector inhibited, but the costs and burdens to business are kept unnecessarily high.

4. **Support for small & medium enterprises (SMEs)**

According to the EU, “SMEs represent 99% of all businesses in the EU. In the past five years, they have created around 85% of new jobs and provided two-thirds of the total private sector employment in the EU.”

Yet, their access to data about the companies they supply, buy from, or compete with is very poor compared with their larger counterparts, who have access to expensive traditional proprietary services. As a result, they lack good due diligence processes, increase their risk and miss out on commercial opportunities. Making company

³ Much of this was highlighted by a report earlier this year on due diligence in supply chains, which listed multiple EU initiatives and directives in this area

<https://op.europa.eu/en/publication-detail/-/publication/8ba0a8fd-4c83-11ea-b8b7-01aa75ed71a1/language-en> “Some suppliers are much ‘bigger’ than their customers, even if those are multinational companies. Suppliers may refuse to cooperate, to respond to investigations or audit requests”

⁴ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0608:REV1:EN:PDF#page=2>

⁵ See, for example, [the Rolling Plan for ICT standardisation: Fintech and Regtech](#), “Business operations benefit from [fintech’s] standardised solutions to support and foster a more competitive and innovative European financial sector...Access to standardised data is essential to perform supervision of financial institutions, monitoring of systemic risk, and market oversight and ensure orderly markets, financial stability, investor protection and fair competition.”

registration data freely available as open data not only reduces the financial costs for such companies, but also, gives them greater ability to scale and compete with larger competitors.

5. Tax

Tax is an emotive subject in almost every EU member state, and perceived injustices in the tax system, and a number of papers⁶ have linked corporate tax injustices to populism. Yet at the heart of corporate taxation is the concept of distinct legal entities and the relationships between them – in the EU this information is actually already public, albeit siloed at national level in official public records at company registers, as was highlighted by a report by the Tax Justice Network last year: “Corporate registries play an important role in preventing corporations from being misused to conceal money laundering, tax evasion and corruption”⁷. A measure of how much EU company registers consider the role their data can play in this area is perhaps indicated by the fact that according to the report only seven responded to the survey, despite repeated requests.

6. The EU Data strategy⁸

Introducing the strategy earlier this year, the President of the Commission, Ursula von der Leyen, said: “Today we are presenting our ambition to shape Europe’s digital future.... I want that digital Europe reflects the best of Europe – open, fair, diverse, democratic, and confident.” A key part of this is the effective implementation of the Open Data and PSI Directive.

7. Corruption and the rule of law

A key fundamental tenet of the EU is the rule of law. Yet this is being threatened through a number of vectors – inequality in tax, particularly around corporations, unequal application of the law, lobbying and the revolving door⁹, and of course corruption. This was highlighted by a Eurobarometer survey last year¹⁰, with 89% of all

⁶ E.g.

https://www.researchgate.net/publication/317747445_Taxing_the_powerful_the_rise_of_populism_and_the_crisis_in_Europe_the_case_for_the_EU_Common_Consolidated_Corporate_Tax_Base and http://eprints.lse.ac.uk/86880/7/Cox_Rise%20of%20populism%20published_2018.pdf

⁷ According to a [report published last year by the Tax Justice Network](#), “corporate registries play an important role in preventing corporations from being misused to conceal money laundering, tax evasion and corruption”

⁸ <https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy> The recently published European strategy for data highlights the structural and system issues that the current regime poses, particularly in High Value Datasets, such as official company register data, for example “High-value datasets are often not available under the same conditions across the EU to the detriment of the use of the data by SMEs that cannot afford this fragmentation”

⁹ Ombudsman slams EU bank watchdog for ‘revolving doors’

<https://euobserver.com/institutional/148325>

¹⁰

<https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/survey/getsurveydetail/instruments/special/surveyky/2235>

respondents saying that action on corruption needs improvement in their country. As the World Bank Puppet Masters¹¹ report illustrated, most large-scale corruption uses companies as the vehicle for the crime. The importance of company register data in this area – and the lack of access to it – was highlighted by a report by the Organised Crime and Corruption Reporting Project and Access Info Europe¹². It is interesting to note that many of the worst scoring countries in this report also scored badly in the Eurobarometer survey.

8. Innovation

There is understandably a significant focus within the EU on enabling innovation, particularly in the area of technology, with billions of Euros spent via grant programs such as Horizon 2020. Companies operating in these areas, however, are almost by definition cross-border. For them, the barriers of finding and researching suppliers, customers and competitors are particularly problematic, as any friction will significantly impede their velocity and add additional strains to what already are high-risk and high-difficulty areas.

9. Financial crime

Financial crime is a serious problem in Europe, highlighted by the Danske Bank scandal, in which an estimated €200 billion was laundered through the Estonian branch of Danske Bank. A failure of Danske Bank to join the dots between entities in different European jurisdictions is central to this whole scandal; similarly it is by using official company register data, investigators have been able to follow the money, and show how those dots could and should have been connected¹³.

10. Standardisation

At the heart of the single market is harmonisation¹⁴, in law and in other areas too, including technical standards. These are critical for the single market, as the extent to which the market can operate effectively is related to the friction to operating within it, or getting information about the market. Standardised company data is critical to this, as it is the dataset of the key participants in the market.

Unfortunately, the dataset of EU companies is not available as a single dataset, nor is it standardised, despite a number of EU initiatives, such as the Core Business

¹¹ <https://openknowledge.worldbank.org/handle/10986/2363>

¹² In [the study](#), Access Info Europe and the Organised Crime and Corruption Reporting Project (OCCRP) tried to obtain full access to the corporate registries of 32 European countries, sometimes using freedom of information laws. They identified 10 obstacles to information access which included high costs, access granted only to nationals and the ability to only search the registry record by record

¹³ A clear example of this is the investigation carried out by Graham Barrow using OpenCorporates data, and detailed at

<https://grahambarrow.com/a-day-in-the-life-of-an-amateur-money-laundering-investigator/>

¹⁴ https://en.wikipedia.org/wiki/Harmonisation_of_law

Vocabulary¹⁵. In part, we believe, this lack of standardization has been down to two key reasons.

First is the lack of a push factor: many company registers appear to prefer the siloed, restrictive approach whereby access to data is limited to small select groups, usually filtered by charging, and hence interconnection (the driver for standardisation) is highly limited.

Second, the lack of a pull factor: effective standards are always driven by an active and demanding user base, who benefit from the reduced costs and increased utility that standards bring. However, because the dataset of European company data is not available as open data, there are just a handful of largely legacy consumers of this data. These legacy companies will have invested time and money coming up with their own internal standardisation approaches, and thus have nothing to gain from it. This benefits them, but increases costs and acts as a barrier to new entrants.

11. EU company law

There are a significant number of EU directives that specifically target information collected on companies for a statutory purpose. However, few consider the mechanisms for the reuse of that data, at least not from the perspective of our increasingly automated, data-driven world. Instead they assume that information today is only needed on a record-by-record basis, rather than the structured data that powers insight, innovation and applications today.

The result is an increasing asymmetry between those creating and running companies, and those seeking to understand them, whether they be journalists, civil society, investors, partners or regulators¹⁶. This has serious problems relating to the rule of law, allowing easy use of companies by organised crime, money launderers, and fraudsters, who use the cover of opacity to act with a high degree of impunity. New regulations brought in last year mean that it will be easier than ever to automatically create companies by computer program, leading to so-called Firefly companies¹⁷ – those in existence for a very short window – just enough for them to execute on their purpose.

12. Trade credit

There are a number of known issues in the trade credit sector, from antiquated credit models, concentration among a small number of players (3 companies are believed to

¹⁵ <https://joinup.ec.europa.eu/solution/e-government-core-vocabularies/core-business-vocabulary>

¹⁶ An example are the [changes to EU Company Law last year, around the digital tools and processes](#).

¹⁷ See Fireflies and Algorithms:

<https://medium.com/@opencorporates/fireflies-and-algorithms-the-coming-explosion-of-companies-9d53cdb8738f>

have 85% of the short-term credit insurance market¹⁸), asymmetries of information, and high costs that have a significant impact on the availability and quality of credit information, on access to trade finance, and on innovation.

At the base of all these issues is the restricted access to company information, as data, with, for example, one company dominating credit information in Italy, with 50% of the market, amounting to two and a half times that of their nearest competitor¹⁹. Needless to say, even a tiny improvement in access to credit, as a result of better information, leads to huge financial benefits.

13. Lobbying & declarations of financial interests

Lobbying in the EU is big business, and the past 10 years have seen many improvements in transparency, at least at the EU level. However, from a practical perspective, even something like the Transparency Register, which is relatively well structured and available as open data²⁰ (neither of these conditions apply to national registers), is essentially a siloed disconnected dataset.

Want to know more about the company concerned from its incorporation documents? This is generally not possible, as there is no identifier for the company. It is not even a requirement to list its legal name or to say where it is incorporated²¹. There are similar problems for elected representatives' disclosures of interests in companies, whether as consultants or shareholders, as these are not explicitly linked to the companies concerned. Nor can civil society solve this, for example by matching to potential companies, as they don't have access to the underlying dataset of EU companies.

14. Non-financial reporting, including environmental impact, pay levels, human trafficking etc

The EU is currently consulting on a revision to the Non-Financial Reporting Directive²², and as the consultation document says, "Non-financial information currently disclosed

¹⁸ According to a [report](#) co-authored by DG Competition, Study on short-term trade finance and credit insurance in the European Union

¹⁹ One Italian company, for example, Cerved, dominates both credit information and company information in Italy, with [by its own estimates](#) 50% of the credit information market to corporate clients, more than 2.5x that of its nearest competitor. This dominance is also based on its historic dominance of the business information market in Italy, which is one of the most restricted registers in the EU, scoring 0/100 on this survey

²⁰ <https://data.europa.eu/euodp/en/data/dataset/transparency-register>, Euler Hermes, Atradius and Coface now dominate the ST credit insurance world and have been estimated to have a combined global share of 85% of ST credit insurance premiums.

²¹ Take for example the listing for "Google" at <https://ec.europa.eu/transparencyregister/public/consultation/displaylobbyist.do?id=03181945560-59>. It's not clear which company this relates to. Is it [Google LLC](#), [Alphabet Inc](#) (the parent of Google Inc), or even [Google Belgium](#) (which matches the address)? Without knowing this, it is difficult to understand the reported data, and identify [underreporting](#), for example

²²<https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12129-Revision-of-Non-Financial-Reporting-Directive/public-consultation>

by companies does not adequately meet the needs of the intended users. Reported non-financial information is not sufficiently comparable or reliable. Some companies from which... users want non-financial information do not report such information. It is hard... to find non-financial information even when it is reported.”

This consultation, and the recognition of the problems, is to be welcomed, as many of the biggest issues facing the EU require companies to act as positive actors in these issues, and be seen to be doing so by their stakeholders.

Clearly it is essential that the information needs to be publicly available to all as data – otherwise it will be largely invisible to key stakeholders, particularly consumers and trading partners. More than this, however, it needs to be connected to the official information about such companies held in registries.

15. **Disaster response, as highlighted by the Covid-19 pandemic**

Any disaster, from environmental to health to economic is going to significantly affect companies, and require their help to tackle it. While this may seem fairly straightforward on the surface, contacting them using tax or procurement records, in practice this is undermined by siloed datasets, unlinked to official legal entity data, particularly for those companies incorporated or controlled outside the territory.

More than this, we are seeing an unprecedented transfer of cash from member states to companies, as a pragmatic way to keep the economy running. Yet while employees and economies are benefiting from this, so are those owning and running companies. It will be important that there is seen to be equity in who subsequently pays for the crisis recovery, otherwise we will see populism once again step in to fill this inequity.

The EU Open Data and PSI Directive

On July 16 2019, the Directive on Open Data and the Re-use of Public Sector Information²³ came into force. Among the changes were moves to increase the availability of official public data (public sector information, or PSI) as open data, through a number of measures.

Most noticeably, as far as access to company data goes, it listed six so-called High Value Datasets that must be made available as open data, including ‘companies and company ownership’. This is truly transformative, and would, if fully implemented, transform EU member states from an average score of 40 out of 100 on this survey, to one of between 80 and 100, due to the depth of the data in general in EU registers.

Unfortunately, what the directive actually means in practice has been left to an implementation phase – a process known as comitology²⁴, in which the details of how

²³ <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024>

²⁴ https://ec.europa.eu/info/implementing-and-delegated-acts/comitology_en

directives are implemented are thrashed out by member states, with the help of the European Commission. This process is at time of writing ongoing, with inputs from a consultation process, in which key stakeholders have the opportunity to make submissions, including via focus groups.

The results of all this – into which OpenCorporates has fed extensively – will not be known until the recommendations are published later this year, probably in September, but it appears there is pressure from a number of countries to limit the company data to a subset of what is collected and published publicly by company registers.

There are also drives to add additional restrictions on use of the data (even though this appears to be prohibited by the directive), to limit the utility of the data, or to add GDPR-like restrictions on the data, even though this would actually be counter to the idea of the GDPR – which is designed to avoid this sort of duplicative piecemeal approach.

The company ownership data referred to in the directive is critical, but also vague, particularly in relation to another key piece of European legislation – the 5th Anti-Money-Laundering Directive (AMLD5). Who ‘owns’ a company is critical in almost every walk of life: who do you actually work for; who are you buying from; who is benefiting from grants and bailouts? These are very real and present concerns. But it’s also critical in making sure that companies can’t be used for criminal and illicit purposes.

This then is the context for ‘company ownership’ – so what does this mean? Business and civil society consider this to be the people and organisations that control and benefit from companies²⁵. In most cases this is the shareholders, the so-called legal ownership.

However, in some companies, particularly those used for money laundering, organised crime or complex tax planning, control and benefit typically happens via complex corporate networks of different types of companies in multiple countries, even using side agreements to ensure control and benefit without actually owning shares.

To find out who really controls the company you need to know the so-called ultimate beneficial owners – shareholding without the obfuscation, if you like. So to understand both who really controls and benefits from companies, and to understand corporate structures of Europe’s biggest structures, both beneficial and legal ownership data is needed.

²⁵ In discussions on company ownership relating the the implementation of the PSI Directive, both business groups and civil society were unanimous in expressing the need that company ownership must mean both legal ownership and beneficial ownership.

The results: overview

The last report on access to company data in the EU²⁶ was published by OpenCorporates in 2012, when the average score was 23 out of 100. Since then there have been some significant changes, both in the EU, and globally. A year after that report, the EU published its open data strategy²⁷, and there have been numerous directives and reports addressing company data. Yet despite this, the average has increased to just 40 out of 100 (jurisdictions score minimum 70 points for publishing a dataset of basic company information under an open licence).

The new Open Data & PSI Directive should, however, change this massively, ensuring all jurisdictions publish company information, including ownership, as open data by July 2021. Providing this happens in full, it would mean the average score will likely be over 90 out of 100, which would truly be world-leading²⁸. However, we understand there has been considerable pushback from some countries, attempting to reduce the effective scope of the directive, or delay it (see below), and so this transformation is by no means certain.

What good looks like

No EU country scores 100 points, nor even 90 out of 100. The two highest scoring countries were until recently the UK (which scored 90) and Denmark (which also previously scored 90), but Denmark no longer makes available a bulk dataset downloadable under an explicitly open licence, reducing its score. It does have a free API, with all categories of data available, potentially scoring 100, but the licence, while largely permissive does not qualify as an open licence, and signing up for an API is not an automatic process.

Rank	State	Score (out of 100)
1	Cyprus	80
1	France	80
3	Bulgaria	75
3	Denmark	75
5	Finland	70
5	Ireland	70
5	Latvia	70
8	Romania	60
9	Belgium	55
10	Slovenia	50

²⁶ https://opencorporates.com/downloads/eu_company_data_report.pdf

²⁷ <http://www.iprhelpdesk.eu/news/eu-commissions-open-data-strategy>

²⁸ On the basis that all company data held by registers would be published as open data, including shareholding and beneficial ownership. Of the 100 points achievable, 10 points are for financial accounts information, which not all jurisdictions collect as structured data

The UK is no longer part of the EU, and thus is not explicitly included in this report, but it does remain a benchmark for what good looks like. This is not just because of the fact that it makes all its data available for free as data, including directors, beneficial owners, filings and accounts documents. On top of this it proactively encourages reuse of its data and backs this up with an active forum for developers, on which Companies House staff answer user questions, announce new features and data, and are receptive to any data quality issues that are raised.

There is also the opportunity for ordinary members of the public to report data quality problems on the Companies House website, which together with active use of their data by tens of thousands of users, provides a powerful data-quality feedback loop.

Finally, there is regular engagement with key stakeholders such as transparency NGOs, and reuse of its data is considered to be a core KPI. Together this has transformed use of the company register data, increasing usage massively, by over 20 times in the past four years²⁹. It's worth noting that this transparency hasn't reduced incorporations, which were over 600,000 in 2019.

What bad looks like

Rank	State	Score (out of 100)
21	Netherlands	20
21	Portugal	20
23	Germany	15
24	Italy	10
26	Austria	0
26	Spain	0

Spain and Italy are two of the largest economies in the EU, and thus are standouts among the member states for their poor scores. Neither makes any data available as open data, and both charge for even the most basic data records on their web portals. Rather than prioritising transparency of company information, they appear to see their key role as revenue generation through restricting access to it.

Special mention in this section should go to the Kamer van Koophandel (KvK) – the Netherlands register. This is an important register not just for the Netherlands but the whole

²⁹ The [2014/15 annual accounts](#) reported the register was accessed 291m times; the [2018/19 accounts](#) reported that this had increased to 6.8 billion times. While figures for API access and open data file downloads aren't listed in the accounts, as of about a year ago, there were over 12,000 API users, and the open data files were downloaded in excess of 600,000 times

of Europe, as it is one of the largest registers in Europe, with a far greater number of companies per capita than, say, Germany, perhaps due to its use by large corporations in tax planning³⁰.

What distinguishes the KvK is that it actively appears to be against open data³¹, and appears to be working to counteract moves towards openness. For example, just over a year before it is due to make the company register and company ownership available as open data via the Open Data and PSI Directive, it is claiming database rights³² on the register, and is proposing using these to prevent other users from storing copies of the data, or making the data available to third parties³³.

This is entirely in contradiction to the Directive, which states, “the high-value datasets [including company register] should be made available for re-use with minimal legal restrictions and free of charge”. Whether the company register would be considered by courts to be able to have database rights is open to question, and it would have serious societal implications if it did. But the directive states further that such rights must not be used to “prevent re-use or to restrict the re-use of existing documents beyond the limits provided for in this Directive”.

Another question is whether the KvK – a private entity³⁴ – should be allowed to set the agenda like this. Unfortunately, the organisation is largely opaque, with accounts which list only gross figures for income and expenditure, and no explanation of how it spends its budget, which is twice the size of UK Companies House³⁵, despite having fewer than half the number of companies in the register³⁶.

³⁰ A [recent report by the Tax Justice Network](#) singled out the Netherlands as enabling large corporations to avoid tax: “It plays a key role in the profit shifting strategies of multinational corporations. In the case of US multinationals, \$70 billion of profit are booked in the country, with just \$3.4 billion paid in tax – an effective tax rate of 4.9%...an amount larger than the total of profits booked in the entire European Union, excluding the other leading corporate tax havens of Ireland and Luxembourg”

³¹ Its [latest annual report](#) says, “Zowel open data als privacy hebben de potentie om KVK (financieel) negatief te raken” (*Both open data and privacy have the potential to hit the Chamber of Commerce (financially) negatively*)

³² https://en.wikipedia.org/wiki/Database_right

³³ In a communication with existing large-scale data customers that has been seen by OpenCorporates, KvK proposed to use the database rights (which we know of no other company register in Europe claiming) to strengthen its revenue stream, and inhibit reuse. Civil society groups have tried to engage with the KvK on these and other issues, but without success

³⁴ The Dutch government describes the KvK as a “public service provider that helps business owners run their businesses successfully” at <https://business.gov.nl/partners/netherlands-chamber-of-commerce/>

³⁵ [The latest KvK annual report](#) shows costs of €200m compared with £71.7m (approx €82.5m) for [Companies House in 2018-19](#)

³⁶ It is not possible, for example, to see the costs and income of running the register, compared with its wider activities of promoting company activity

How does this compare with other countries?

The EU's average score of 40 is by no means the worst in the world – the average for the US is 31/100 (in the US companies are incorporated at the state level). However, unlike the EU, the US has no states that score 0 out of 100 (the EU has two: Spain and Austria, and one with just 10/100: Italy). Since the UK left the EU, the average score has dropped, but only by a little – as while the UK scores 90/100, its departure also saw the departure of Gibraltar, which scored 0/100.

A tale of two states: Sweden vs Denmark

The uneven nature of access to official company register data in the EU can be seen in two countries right next to each other: Denmark and Sweden. Denmark scores a respectable 75 out of 100 – not the best in the EU but light years ahead of Sweden. All of the register is available for free, and there is a free API that allows all the data, including shareholders and accounts, to be downloaded for free as structured data.

Sweden on the other hand currently publishes no open data on companies, and in fact even restricts all but the name, company identifier and status to paying users of the website. As such, it scores a very poor 25 out of 100 – 20 for making that basic data available for free, and 5 for terms of use that may not restrict usage but nevertheless don't provide the legal certainty of an open licence.

Other notable issues relating to EU company data

Publishing frequency

There are a number of jurisdictions which publish company data infrequently. Starting from this report, we consider this to be a data quality issue and are down-scoring some countries accordingly (data quality issues result in a reduction of 10 points). Romania, for example, publishes its open dataset only every 3 months. Bulgaria has failed to update its data for over a year, despite being contacted a number of times. After communicating with the register, we have been told that they will be publishing the dataset this month (June 2020), and will be publishing regularly after this. Given these assurances, we are scoring Bulgaria as if they are continuing to publish, but with the data quality deduction. We will check that this is done at the end of June 2020, and rescore should it not be the case.

Licences

Company register data isn't important only in its own right. It's a foundational dataset – the basis on which all other data related to business, our digital lives, and increasingly society itself depends. For this reason, it is important for the data to have no restrictions on use, as otherwise it will undermine the utility of it. For this reason we recommend the [CC0 public](#)

[domain licence](#). Licences that require attribution appear to be superficially attractive, but in practice bring no public benefit, and add real challenges to utility³⁷.

Share-Alike³⁸ requirements, whereby the reuser is required to make the data available under the same licence, are highly problematic from a number of perspectives, and they are in practice usually a way of allowing the data still to be charged for. This is because a share-alike licence means that the data cannot be used commercially, and this means that there can be no income stream to sustain any work using the data.

So while one-off quick uses – say downloading the dataset and analysing it – are still possible, longer term transparency and public interest uses are not, and nor does it do anything to support innovative use cases, nor to reduce the costs to SMEs. Due to these problems, share-alike licences score only 5 points in the survey (compared with 30 points for CC0 or attribution-only licences).

An example is Estonia, where the open data is made available under a share-alike licence, but to access it without this you must pay €286,000³⁹.

Identifiers

Ask most people how they would identify a company and they would say by the name. But company names are highly problematic as identifiers, for two reasons. First, they change surprisingly frequently – not only can the same company have many different names over its life, the same company name can apply to multiple different and unconnected legal entities over a period of time. Second, they can present all sorts of issues in how they are represented (MICROSOFT RESEARCH & DEVELOPMENT FRANCE SAS vs Microsoft Research And Development France Société Par Actions Simplifiée, etc).

³⁷ When one is dealing with facts (this company was incorporated on that date; this person is director of the company), attribution is not required – you wouldn't think to give attribution if you named the Minister of Finance for Germany, for example, even if you got that fact from a governmental database of public officials. When you start to combine data together, you get complex [provenance chains](#), lists of attribution that can often be much bigger than the data itself. In addition, it is not clear when attribution is required – yes, when the data is used directly, but what about if only part of a datapoint is used; what if it is used in inferences, or to train machine-learning systems? This all adds legal uncertainty for little benefit.

Finally, it's not clear what benefit the register or government gets from requiring attribution – no-one doubts that it is the official source. In fact the single thing that undermines this most is only making the data available to proprietary vendors, who have the ability and incentive to not attribute the data. Our experience with open data from company registers has been that in practice, attribution is often given even when there is no requirement to do so, as it adds authority to the statement and context for the user, but this should be optional rather than a requirement

³⁸ <https://en.wikipedia.org/wiki/Share-alike>

³⁹ <https://www.access-info.org/blog/2015/07/02/business-transparency-infographic-3/>

These problems are magnified when combining datasets together, when trying to understand for example, who are the biggest suppliers to government in the EU, or recipients of EU grants, etc.

As a result, clearly identifying companies has become increasingly recognised as critical for good, trusted and useful data – using well-defined and well-formed *public* identifiers, separate from the name and any transient attributes. This removes ambiguity, legal uncertainty, and streamlines processes. It is particularly important to allow clear and unambiguous identification of a company in other datasets, for example in the EU’s lobbying register, or in government procurement data.

Perhaps surprisingly, there has been very little work done in the EU on company identifiers. In some countries, the identifier for the company is enshrined in law and must be used in all correspondence⁴⁰. In some, the form of the identifier is described in regulations and publicly described⁴¹, to bring clarity to it. In some countries the company identifier actually is the same as the tax number.

Unfortunately, not only does the usage of the company identifier vary from country to country, so does the quality, and availability, with significant issues in some of the largest member states. For example:

- In Italy for example, the company identifier is not available in the free search of the company register, presenting various obvious issues.
- In Germany, company identifiers are not well defined, being a mix of the name of the district court that identified it, the type of register, and a local number. How these are combined and represented varies from situation to situation, and, worse, when a company moves from one court area to another it gets given a new identifier.
- In Spain, it’s actually not clear what should be used as the identifier. There is a complex identifier issued by the local chamber of incorporation representing the entry in the local register, but this is universally ignored in favour of the CIF, the company’s tax number (which is also the primary identifier in the central company register).

One option that has been discussed is to use the so-called LEI (Legal Entity Identifier) as a universal identifier for EU companies. The Global Legal Entity Identifier System⁴² was set up by the G20 to address the problem of identifying legal entities in financial markets, and is increasingly being used in other fields too. As well as being permanently associated with a single legal entity, LEIs contain no information about attributes (which might change, and with it the identifier) and are open and non-proprietary. In fact a number of the EU company

⁴⁰ E.g. <https://www.gov.uk/running-a-limited-company/signs-stationery-and-promotional-material>

⁴¹ E.g. <https://www.insee.fr/en/metadonnees/definition/c2047>

⁴² <https://www.gleif.org/en/about-lei/gleif-management-of-the-global-lei-system>

registrars are already issuing bodies for the LEI system (so-called LOUs⁴³), and the LEI is mandated for all participants in the financial industry and stock markets by the European regulators of those industries.

There has been movement from the EU in recent years, with the creation of the “EUID”, which was mandated for use in a project to connect company registers. Unfortunately there is very little information on the EUID beyond the original EU document⁴⁴ outlining the technical implementation of the interconnection of registers, and even that is hard to find unless you know what you are looking for⁴⁵.

The EUID unfortunately has a number of problems:

- Its purpose isn’t clear. Is this just for the technical connection of registers, or is it meant to be used more widely? The lack of public-facing documentation would indicate the former.
- There is a lack of awareness and use – it’s not used by companies, nor to the best of our knowledge by governments outside of the interconnection of registers.
- The business rules of the system are not clear. For example:
 - How are the various elements of the EUID identifier made up⁴⁶?
 - Does the EUID change if the register identifier changes⁴⁷?
 - What happens if there are multiple registrations in the same register, as in Germany⁴⁸?
 - What are the code lists which are used in the identifier⁴⁹?
 - How does one know whether the optional verification digits have been used in an EUID, and how they have been calculated?

Many of the problems relating to the identification of companies in the EU (as well as other problems) could be solved – either by moving towards the LEI as the primary identifier (which

⁴³ <https://www.gleif.org/en/about-lei/get-an-lei-find-lei-issuing-organizations>

⁴⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32015R0884>

⁴⁵ A search for EUID returns mostly information about IDs used in operating systems. Only when restricting the search to official EU sites do you find a link to the official document (9th result at time of writing): <https://www.google.com/search?q=EUID+site%3Aeuropa.eu>

⁴⁶ For example, how many characters are the individual elements?

⁴⁷ The implication from the official document is that it should, but this is contradicted by a Hungarian [business consultancy firm’s website](#) about the Hungarian implementation: “A change in a Hungarian company’s registration number does not result in a change in the European Unique Identifier”

⁴⁸ From a brief investigation, it would appear that there is an EUID for each register entry, even though this is not what you would expect from the regulation, which states that “EUID shall be used to unequivocally identify companies”. For example a search of the [eJustice portal](#) that provides a portal to the interconnection service for “AGCO Finance GmbH” yields both the historic Hannover registration as well as the current Dusseldorf one

⁴⁹ For example the codes which represent individual registers

would bring multiple other benefits too), or by truly embracing reuse of the data, to improve quality and work in a collaborative way to solve these issues.

What is a company (and other data quality and standardisation issues)?

One of the challenges of dealing with EU company data is that the registers perform different roles in different countries. In some jurisdictions, the register actually incorporates the company; in others the company is actually created by courts or local chambers of commerce, and the register merely records the information. Another problem is that what type of entities are recorded in the register varies from country to country too – some include foundations and trusts, some don't; sole traders are recorded in some countries, but not all; some even include government entities and churches.

While this inconsistency can cause issues, far more problematic, however, is the German register, where a record represents a transient registration tied to an entity's location, and a new registration (and thus new record, and new identifier) is created each time a company moves its headquarters⁵⁰. This has two consequences: difficulty in identifying the correct record, and key data about the company is separated across multiple records. For this reason, we have marked this situation as a data quality issue.

GDPR and privacy

No discussion of data in a European context can avoid talking about GDPR (the General Data Protection Regulation), and privacy. GDPR has undoubtedly changed the landscape of data and data use not just in Europe, but worldwide, influencing legislation in other jurisdictions (e.g. the California Consumer Protection Act, or CCPA), as well as changing many aspects of how personal data is stored and used.

However, the GDPR is a new law, and there is very little in the way of case law, and none that we know of that has been ruled on by the European Court of Justice. This means that there is much uncertainty, in particular where the balance of rights will fall between privacy and other rights (the right to free speech, for example). In the CCPA, for example, public records are specifically excluded from the scope, yet it is not clear whether that will turn out to be the case in the EU.

In addition, there are three other aspects which are relevant to this report.

The first, is the tendency to take this complex and subtle legislation and try to reproduce a simplified aspect in other legislation and regulation, even if that regulation is already covered by GDPR. This actually undermines the GDPR regime, and prevents it from working effectively,

⁵⁰ See

<https://blog.opencorporates.com/2019/02/19/wait-what-the-problems-of-company-numbers-in-germany-and-how-were-handling-them/> for more details

and should strongly be resisted⁵¹. Similarly GDPR is used as an excuse for not releasing company information as open data, even though the same data is being sold to third parties apparently without any concerns.

The second is the way it increases the asymmetries of power and agency. For example, the requirements it puts on companies actually strengthen the position of the giant tech companies for whom large teams deciding who should and should not be given the right to be forgotten is just a relatively small cost of doing business.

Another example is the growth of a whole new industry around removing traces of people's past from the internet⁵². Unfortunately this is increasing the asymmetries of power. The vast majority of people are not directors of companies, or beneficial owners, and it is only the relatively rich and powerful (including politicians and criminals) who can afford lawyers to stifle free speech, or remove parts of their past they would rather not be there, from business failures to associations with criminals. OpenCorporates, for example, was threatened with a lawsuit from a member of one of the wealthiest families in Europe for reproducing a gazette notice from the Luxembourg official gazette (a publication that contains public notices). We refused to back down, believing we had a good case in law and in the public interest, and the other side gave up. But such so-called SLAPP⁵³ suits are becoming increasingly common, although unlike many US states there are currently no defences in place to resist these in the EU, despite pressure from civil society to address this⁵⁴.

The third issue is a more subtle one. Whilst it is out of scope for this report, it is nevertheless important to mention: as an unintended consequence, GDPR is having a side effect of strengthening proprietary data owners and providers, both in comparison with SMEs and civil society, but also with ordinary citizens too. This is happening in two ways:

- The processes and legal advice required around GDPR, particularly dealing with complaints, most of which are mistaken, is very capital intensive, requiring significant expenditure on legal and response systems.
- Those organisations whose primary purpose is promoting transparency (NGOs, journalists, social enterprises) are ironically disadvantaged compared with those who are purely commercial and effectively under the radar. Most people are not even aware of such companies, compared with websites devoted to increasing transparency (for example, improving oversight of government procurement, grants or lobbying).

⁵¹ For example, in the Netherlands Beneficial Ownership Register:

<https://openstate.eu/en/2020/06/nederlands-nederland-werpt-onnodige-barrieres-op-tegen-register-dat-vecht-tegen-corrupcie-en-witwassen/>

⁵² For example, two such companies advertise: "If your Reputation is suffering from negative content. We specialise in burial or removal." and "We erase your past on the internet"

⁵³ https://en.wikipedia.org/wiki/Strategic_lawsuit_against_public_participation.

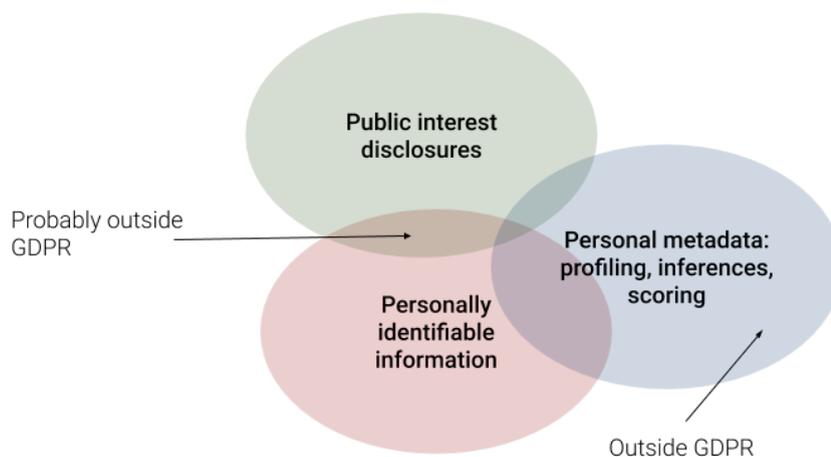
⁵⁴ <https://www.business-humanrights.org/en/27-ngos-call-on-european-commissioner-to-include-all-parties-impacted-by-slapps-in-new-eu-legislation>

There's one other crucial factor – GDPR is not actually about privacy but rather data protection, and many of our most private aspects – how we think, what we like, how we are profiled – are not covered by GDPR. This has been highlighted effectively by Professor Sandra Wachter, who highlights that not only are such aspects not covered by GDPR but that other EU legislation is putting this data further out of reach of us⁵⁵.

With today's technology, that means that we can be profiled, judged, categorised and targeted without anyone knowing or storing our name. This information can then be used to either sell to you, judge your creditworthiness, or serve up political advertising designed to trigger your fears and hopes.

At the same time, the automatic assumption that all Personally Identifiable Information (PII), someone's name for example, is private is highly problematic, confusing both citizens and policy makers, and further undermining democracies and fair societies. As an obvious case, it's critical that we know the names of our elected representatives, and those in positions of power otherwise we would have an opaque society where decisions are made by nameless individuals with opaque agendas and personal interests – such as awarding a contract to their brother's company, for example.

As the diagram below illustrates, there is some personally identifiable information that it's strongly in the public interest to know. Take the director or beneficial owner of a company, for example, of course their details are PII – clearly you need to know their name (and other information too), otherwise what actually do you know about them, or the company? That some unnamed individual has been given special protection under law to be shielded from the company's debts and actions, and yet can benefit from its profits.



⁵⁵<https://www.law.ox.ac.uk/business-law-blog/blog/2018/10/right-reasonable-inferences-re-thinking-data-protection-law-age-big>

It seems likely – and is to be hoped – that we will develop a more nuanced, and deeper, understanding of privacy as awareness of the issues increases, and case law starts to bring clarity to the ambiguities of the GDPR.

Foreign branches: when is a company not a...

Foreign branches are registrations in a country by companies incorporated in another country, usually because they have a physical presence or significant business activity in the ‘foreign’ state, and can be a lighter weight, more cost-effective route of establishing a presence in another country rather than creating a subsidiary legal entity.

Branches, however, presents some interesting issues. For example:

- The company is operating under both the laws of the local jurisdiction, and the home jurisdiction.
- In the case of insolvency, it is the home jurisdiction that prevails, and with a few exceptions (e.g. banks) there are no assets owned locally. Thus much of the external risk lies in the local jurisdiction (e.g. cleanup of a manufacturing plant), but the benefits accrue to the home jurisdiction.
- There are countless data quality issues that OpenCorporates has identified – for example where the branch is marked as live, but the home company has been dissolved, or where there is a mismatch between directors listed in the branch registration and those listed in the home registration.
- It appears that this valuable dataset is little used by the EU member states, even though it has significant potential in the areas of tax collection and regulation.

In theory it is not a hard problem to link branches to the home companies – *if* you have access to all register data in the home jurisdictions. Indeed OpenCorporates has done this for tens of thousands of companies in the EU (and millions in the US, where branches are even more common). Performing this task often highlights a number of data quality issues, or anomalies, especially when the home company dissolves or changes name, but the branch registration does not. It also frequently adds data and insight to both the branch and home company records, either providing links to additional data, or further understanding about the areas in which it is operating.

Black Box data: why it's problematic, and how closed registers enable it

The canonical source of core company information in Europe – which companies exist and existed and what their key attributes are – are the company registers of the member states.

However, if you need it as data (and many companies do), and particularly if you want it as standardised data across multiple countries, you must use intermediaries – legacy proprietary business information companies, who combine this data together and then make it available via platforms or as bulk data dumps. This in itself is no problem – adding value is what business should do.

The problem is that what these legacy proprietary companies produce is so-called Black Box data – data where it is unclear what it represents, where the source is obfuscated, data that uses proprietary identifiers, and data which is limited to a relatively small number of users, and thus has poor data-quality feedback loops.

This was always intrinsically problematic, but today it is no longer fit for purpose. As a recent white paper called [The White Box Data Revolution](#) makes clear, today's data-driven applications need data with a clear and transparent model, data with full provenance, with open identifiers, and that is public, giving the best possible data-quality feedback loops.

So if the data is not what's needed by the market, why does it still exist? The problem is that charging for the data shuts out anybody that doesn't have an existing strong customer base, denying innovators the opportunity to disrupt the flawed status quo, and provide data consumers with what they need. Even if registers charged, say on average €50,000 per year for access to their data (and some charge millions of Euros per year⁵⁶), that would cost anyone wanting to assemble this data over €1.3million a year

Bulk data vs API: which is better?

If the purpose of a register is to contribute to a trusted business environment by making the official public record easily available, then it should be providing the data in the most appropriate form to maximise its usage.

Most jurisdictions that start publishing open data do so by publishing a regular dump of the company data, typically once a month. This is typically very easy, cheap, quick and simple to implement. Depending on the complexity of this information, this may be published as a downloadable CSV file, as a file containing data in JSON format, or using an open data platform, such as CKAN or Socrata.

⁵⁶ Access Info Europe attempted to find out how much it cost to buy the company register data in bulk, with only partial success, but those figures together with ones obtained by OpenCorporates indicate that the cost of assembling the official company register data is in the many millions of Euros per year. Not surprisingly there are only a very small number of companies that buy all of this data, causing huge market inefficiencies.

Later on, open data registers will often create an open API (a way of computers transferring data to other computers over the web), allowing direct queries of the register. This means the information they are getting and basing their decisions on, or presenting to users, is as current as can be.

Both forms of access have their pros and cons, and both are needed for the register to perform its role of providing the public record to help create a trusted business environment.

The principal downsides to the monthly bulk data approach is that as soon as it is published it is out-of-date (assuming changes are constantly happening), and that you need a system that can hold all the records in the dataset, which may be several million. For many current use cases the latency does not matter, but as we move to a real-time data world (see below), it may become increasingly problematic.

The main problems with an API-only approach is that many use cases require access to the entire dataset in one go, e.g. analysis and data science. While it is possible to build up the entire dataset using hundreds of thousands of API calls, it is inefficient for both publisher and user.

Bulk data use cases

Bulk downloads are the most basic form of data publication and are critical for most use cases, including:

- **Combining datasets**
Combining with other datasets, e.g. company records from other jurisdictions, or with business licences, procurement data, sanctions lists, etc.
- **Data analysis**
Analysing the dataset, e.g. identifying trends, anomalies, creating models.
- **Foundational data**
Use as foundational data, e.g. so that the official entity record is the basic datapoint to which the user's other dataset is appended.
- **Speed & simplicity**
Use cases where speed and low technical barriers to entry are important. Many organisations have the ability to deal with a single dataset – whereas writing the code to interface with an API and make the information usable may take skills and resources that the organisation simply doesn't have (e.g. NGOs, journalists, law enforcement or regulators).
- **Security**
Many government and law enforcement agencies, as well as some large financial

institutions, are restricted in their use of external APIs. They are bound by rules that, for example, prohibit them from making queries about entities they are investigating.

API use cases

Though web-based APIs have been around for many years, they are still the exception in open data. However, for a number of use cases they are critical, including:

- **Single queries**
For example in due diligence.
- **Low-latency queries**
In some cases, it's critical to have access to very up-to-date data. An example of this is where regulations require it (e.g. anti-money laundering or know your customer etc), or where the old data will give the wrong result (e.g. a newly formed company opening a bank account).
- **Where smaller record numbers are needed**
Use cases where importing the entire dataset is problematic, or not cost-effective. If a user needs only a small number of records (e.g. in an onboarding process), then downloading and importing millions of records is inefficient and will introduce unnecessary technical problems.

Keeping bulk data fresh

As we move to a real-time corporate data world (see below) it will be critical to be able to keep up-to-date with the official source. The best APIs support the retrieval of records that have changed in a certain period (e.g. the past 48 hours). This allows bulk data to be kept up to date by being refreshed by API calls, bringing the best of both worlds to users.

The real-time, computer-controlled future of company formation

Many of us acknowledge that we now live in a data-centric world. Data powers almost every part of our lives – including communications, friendships, work, shopping, navigation and media. However, what's been less well acknowledged is that we are moving into a real-time world. How we consume news, buy music, change routes or make friends instantaneously is evidence of this.

Of course, there are still processes which are done manually, with heavy friction and sometimes for a good reason (e.g. the judicial system, drawing up laws, voting for high office). Others are done manually for historical reasons, and these will rapidly disappear with advances in technology. A recent example is language translation. A much older one is stock market trading. This was once done manually, with paper and phone. Now it is done almost entirely electronically, and with that move we saw first algorithmic trading, then programmatic

trading and finally high-frequency trading, such that the majority of trading on stocks is now done automatically by computers.

A similar change is coming to company formation, as companies are incorporated, dissolved and maintained automatically. As the essay [Fireflies and Algorithms](#) makes clear, this future is already here in places, with a significant number of registers including the UK already allowing this, and more to come (with changes in the EU's Company Directive).

This – together with the increasing demands of numerous use cases, from tech to organised crime – will lead to a massive increase in the speed and complexity of companies, with companies being created for specific deals (this already is quite common), and lasting for days, minutes or even seconds.

In this context, having access to data on a monthly basis will be like doing historical research. It's interesting, but not what you need for many cases, and will seem as quaint as stock market data that's days old. We believe that within the next 5 years, the base standard for open data – particularly in dynamic areas such as company data – will be combined bulk data export and APIs that allow users to keep their data fully in sync with the official record. This is already happening in the most advanced jurisdictions⁵⁷.

Why charging for company register data is so problematic

Company registers are public registers of companies that have been incorporated or registered in a jurisdiction. Without easy access to this information, it's not possible to know whether a company is still active⁵⁸, whether a company is being impersonated, or even whether it has ever existed at all.

These questions have always needed answering, but today's world is very different to the one 20 or 30 years ago. Today, much of business is cross-jurisdiction and the role of companies in our lives has massively increased. In short, we all need access to company information, and we need it when we want it and how we want it.

This means that it must be available as structured data, so that it can be queried, analysed, combined with other datasets, and used in the myriad of applications that we routinely depend on in work and play.

Of course, it is possible for registers to charge users for this public information – and maybe 20 or 30 years ago, when only a handful of credit reference agencies wanted the data, it was justifiable. However today, we all need access to this data – whether we are a small business,

⁵⁷ For example Washington State has both bulk data dumps and an API, and the UK company register is trialling a so-called 'streaming' or 'push' API, which pushes changes to users as soon as they hit the register.

⁵⁸ For example, see:

<https://www.eveningexpress.co.uk/news/business/firm-at-centre-of-lekoil-scandal-struck-off/>

a journalist, an out-of-jurisdiction law-enforcement agency, an NGO, or just a citizen, and whether we use the data directly (in the UK last year there were over half a million downloads of the company register’s open datasets), incorporated in other data by a FinTech company, or via a platform such as OpenCorporates (which has millions of users a month). In this context, to use fees as a way of restricting access is highly problematic, and against the public interest.

It also makes no sense economically. Because the main users of company register data are businesses, by charging for the data you are essentially taxing businesses, and in the process adding friction and overheads to their use of the data – distracting them from their main activity. In addition, if the companies you are selling to are a small number of legacy data aggregators, you are also propping up their out-of-date uncompetitive business model, stifling innovation, and adding further costs and further friction to business (for example the margins charged by the aggregators or the restrictive licences).

Economically it makes far more sense to cover the costs of the register via filing fees. Companies that incorporate are the ones getting the benefit of incorporation. Filing costs are a minor cost for business, and the net cost to business would actually be reduced, as they would no longer have to bear the costs of the extra work and friction that charging brings.

Finally, it is worth stressing that central governmental portals are no silver bullet either. However good their intentions, governments do not have a good track record in building websites that users want to use, and should not try to crowd out innovation and user need by doing so⁵⁹. Instead they should act as the authoritative source, much like a tax authority or legislation website provides the canonical source of tax rules or legislation, with other sites providing interpretation and insight.

Benefits open company data provides for EU countries

Although this report focuses on the benefits to users of making company register data more accessible, there are considerable benefits to the member states themselves, and to Europe as a whole. There are two types of benefit we have identified: direct (i.e. felt directly by the state) and indirect (benefits for others, but that would indirectly benefit the state).

Direct benefits to EU member states

- **Improved data quality**

By increasing the number of users and use cases, the registers would see significant

⁵⁹ The central portal for access to the data made available to the data made available under the BRITE directive is a case in point, and appears to be little used. We were not able to find traffic statistics anywhere for the portal, but estimates by web statistics services estimate that the whole of the e-justice.europa.eu domain (which includes this portal, as well as many other sections) attracts a fraction of the traffic of OpenCorporates, for example

improvements in data quality, due to better feedback loops. OpenCorporates itself receives countless numbers of reports of state registers that are out-of-date, or for example where people have been erroneously listed as directors.

- **Reduced costs for data acquisition**

As well as publishing company information, the countries and the European institutions are also significant consumers of it too. Currently these costs are inflated by a number of factors, including:

- A small number of suppliers due to the difficulty of accessing the data, meaning little competition.
- Significant markup of fees paid to registers for acquiring the data.
- Restrictive licences that prevent key use cases being addressed.

Publishing company register data as open data addresses all these problems, reducing costs, allowing new suppliers into the market (or allowing the governments to consume the data directly) with lower costs and innovative solutions.

- **Reduced costs when using the data**

While the proprietary Black Box data that states are currently forced to buy appears to make life simpler, there are multiple problems that introduce complexity, lock-in and costs when using data:

- It is nearly always disconnected from the official record of the legal entity, making it unsuitable for many use cases, especially those that may end up in court, as only facts can be used as evidence (e.g. company filings), not Black Box proprietary data.
- In many proprietary Black Box datasets the records don't even represent companies, but buildings, addresses or a vague concept of 'business'. This is highly problematic in regulation, licensing and taxation – increasing costs, effort and impeding the effectiveness of the agency.

- **Improved ability to recover missing taxes**

As well as increasing costs unnecessarily, the current lack of access to data, and the silos that result, mean that member states are missing out on tax revenue.

- **Better law enforcement and regulation**

Law enforcement agencies are increasingly using data to perform investigations – helping to identify and track down financial crime, fraud, stolen assets and organised crime – all of which routinely use companies to carry out the crime, or hide the proceeds. Using OSINT (Open Source Intelligence) such as open data from company registers is becoming a routine method for them, allowing them to perform

investigations quickly and effectively.

- **Less benign environment for the illicit use of companies**

No state wants to have the reputation for providing a cosy environment for companies incorporated or operating there to be used for illicit purposes – particularly money laundering, tax fraud and organised crime. Yet this is precisely what they are doing when they restrict access to the underlying data.

- **Enhanced reputation for transparency and innovation**

Company registration is rapidly changing – with the key trends being digital identities, low latency, data quality and particularly open data access. Those registers that aren't innovating will increasingly be left behind.

- **Better procurement data**

By providing core company data as open data, government procurement could be considerably improved, with consequent improvements in transparency, efficiency and decision-making. Procurement software vendors (governments at all levels, from national, down to local tend to use standardised procurement systems) could easily integrate official company information in their systems, not only improving utility but also data quality too.

Indirect benefits to EU member states

The reason company registers exist is to provide a public record of companies that are incorporated or operating in a jurisdiction. There are multiple indirect benefits to the jurisdiction to making this available as high-quality, fresh open data, including:

- **More efficient business environment**

Opening up the data means companies can more easily access and use the data in their country (where much of their business will be), either directly, or through third-party providers.

- **More trusted business environment**

Opening up the data means there is more trust between businesses. All other things being equal, would you rather do business with a company in a jurisdiction where you could easily access the data, or one where it was hidden away?

- **More innovation**

The relatively tiny number of bulk users of EU company register data – often less than 20, virtually all of them legacy data companies – is a huge indicator of the lack of innovation around this data. Compare this with the UK's open data API, which has over 10,000 different users making 7 billion calls a year.

- **Reduced costs for business**

Like the member states, business suffers from both the excess costs and excess friction caused by inefficient markets where public data is effectively only available as limited-use, poor quality proprietary data.

- **Less benign environment for the illicit use of companies**

It's not only the EU countries that suffer from the illicit use of companies. Business does too.

OpenCorporates will be running pilot programs later this year to help governments make best use of their company data in a variety of areas, leveraging OpenCorporates knowledge, data, and expertise. Contact pilots@opencorporates.com to register your interest.

Detailed results

Country	Total (Max 100 points)	Basic data online	Licence	Structured data freely available	Directors	Annual accounts	Shareholders	2017 score
AUSTRIA	0	0	0	0	0	0	0	0
BELGIUM	55	20	30	0	5	0	0	50
BULGARIA⁶⁰	75	20	30	10	10	5	0	85
CROATIA	25	20	0	0	5	0	0	20
CYPRUS	80	20	30	20	10	0	0	30
CZECH REPUBLIC	40	20	5	0	5	5	5	60
DENMARK⁶¹	75	20	5	20	10	10	10	90
ESTONIA	60	20	30	10	0	0	0	60
FINLAND	70	20	30	20	0	0	0	45
FRANCE	80	20	30	20	0	10	0	80
GERMANY⁶²	15	10	5	0	0	0	0	20
GREECE	35	20	5	0	5	5	0	20
HUNGARY	25	20	0	0	0	5	0	25
IRELAND	70	20	30	20	0	0	0	45
ITALY	10	10	0	0	0	0	0	10
LATVIA	70	20	30	20	0	0	0	70
LITHUANIA	25	20	5	0	0	0	0	25
LUXEMBOURG	20	20	0	0	0	0	0	25
MALTA	25	20	0	0	5	0	0	20
NETHERLANDS	20	20	0	0	0	0	0	20
POLAND	30	20	5	0	5	0	0	30
PORTUGAL	20	20	0	0	0	0	0	20
ROMANIA⁶³	60	20	30	10	0	0	0	70
SLOVAK REPUBLIC	40	20	5	0	5	5	5	40

⁶⁰ Bulgarian open data has not been updated for nearly a year. However, following enquiries, the Bulgarian government has assured us that the data will be published in June 2020. Given the issues with publication, we have considered this to be an infrequently published dataset, and marked accordingly

⁶¹ Denmark has retired the data download service, which had a fully open licence. The remaining API service, is 'mostly open', but has additional terms and conditions, specifically relating to marketing

⁶² Multiple records and identifiers per legal entity (see [What Is A Company](#), above)

⁶³ Published roughly quarterly

SLOVENIA	50	20	0	20	5	0	5	45
SPAIN	0	0	0	0	0	0	0	0
SWEDEN	25	20	5	0	0	0	0	50

Methodology

The research for this report was all conducted in February to May 2020, following on from earlier research for the [Open Company Data Index](#). The methodology and scoring is the same as for the Open Company Data Index, with the addition of a down-scoring of the free machine-readable data by 10 points for data that is published less than every month.

Discovery

Finding information about how to access the underlying data from a company register is often not straightforward. Sometimes there are clear links to how to access the data, but these are usually aimed at legacy commercial buyers. Even when a jurisdiction is publishing company register data as open data, this is often not obvious. It may be hosted on a completely different site, and often is not even linked to from the register’s page.

Scoring

Company registers should be public, as they are the public record of artificial entities given legal personality by the state for the benefit of society. In a free and open society, this important information should be free to use and reuse for all, without charge and without restrictive license conditions.

This material should be in the public domain, with no license restrictions at all. For this survey, scoring has been done based on whether the license complies with the generally accepted Open Knowledge Definition.

Additionally, the underlying data should be available to all without fee, and without restrictions, as machine-readable data. If the data is not available freely, it can only be accessed by those with the resources to pay, thus undermining scrutiny, innovation, and a free and open market.

Scores were assessed on the following basis –with a total of a possible 100 points:

1. Unrestricted online search (no cost, no registration, search feature)

Is basic information online and available to search without charge or registration? (20 points)

This is the base minimum threshold for an open company register. In short, if a company register cannot be searched for without charge, restriction or registration (which implies restriction) then the register is essentially closed to the public.

In cases where there are central registers that aggregate local regional registers (as in the case of Spain or Germany), scoring is based on this central register, rather than the regional registers.

Additional factor: reduce by 10 points if the register has known quality problems (e.g. many duplicate records), is not frequently updated, or doesn't give the status of the company (i.e. whether it is still in business) or doesn't provide unique identifiers. Reason to be recorded.

2. Openly licensed

Is there an explicit open licence (e.g. CC-0, UK Open Government Licence)? (30 points)

- 30 points for a licence that conforms to the Open Knowledge Definition, with at most attribution requirements
- 5 points for no explicit licence, or a licence that conforms to the Open Knowledge Definition with Share Alike restrictions
- 0 points for a licence that explicitly prevents reuse or otherwise fails to conform to the Open Knowledge Definition, including catch-all closed licences (e.g. All Rights Reserved)
- Additional data to be captured: URL of terms and conditions and/or to the licence (e.g. UK Open Government Licence)

3. Free machine-readable data

Is the basic information freely available as data, either as a free data dump or via a free API? (20 points)

- 20 points if the data is freely available
- 0 points if there are any restrictions to having the data freely available
- Clarification: an API may require registration, but must add no additional restrictions to the use of the data, nor to charge for access to the data. Download dumps should not require registration or additional restrictions

Additional factor: reduce by 10 points if the data is updated infrequently (i.e. less than every month).

4. Data depth – directors

Does the publicly available information include a list of company directors for each company?

- 10 points if it includes a list of company directors for each company as freely available data
- 0 points if it does not include a list of company directors for each company
- Clarification: the register should list all directors for a company, or at least all executive directors. Where a company register clearly only makes a single director available it should not be considered to meet the threshold

5. Data depth – annual accounts

Does the publicly available information include annual accounts for each company? (10 points)

- 10 points if it includes annual accounts for each company available as freely available data
- 5 points if it includes annual accounts for each company freely available, but not available as data (e.g. only on a web page, or PDF/image documents)
- 0 points if it does not include annual accounts for each company

6. Data depth – shareholdings

Does the publicly available information include significant shareholdings for each company? (10 points)

- 10 points if it includes significant shareholdings for each company available as freely available data
- 5 points if it includes significant shareholdings for each company freely available, but not available as data (e.g. only on a web page, or PDF/image documents)
- 0 points if it does not include significant shareholdings for each company
- Clarification: to qualify, a register must include detailed shareholdings, including the number or percentage of shares held, and must be contemporaneous (i.e. updated as the shareholding changes, not once per year)

About OpenCorporates

OpenCorporates is the largest open database of company information in the world, containing information on over 185 million companies in 130+ jurisdictions, all from primary public sources, and all freely available on the OpenCorporates website.

Founded 10 years ago, it is a public-benefit company, whose primary public-benefit mission is to make company information more accessible, more useful and more usable. It regularly works with governments on data-quality and accessibility, and with journalists and NGOs on investigations, including with the ICIJ on the Panama Papers.

Errors And Corrections

OpenCorporates has a policy of correcting errors as soon as they are brought to our attention. If you believe that any of the scores in this report are incorrect or out-of-date, please contact us at info@opencorporates.com.

OpenCorporates would like to thank The Mohn Westlake Foundation for supporting the development of this report.