opencorporates

US Company Data: State of the Nation 2020

How accessible is official company register data in the US?

OpenCorporates Ltd is part of the OpenCorporates group, a public-benefit enterprise whose goal is to create, increase and promote transparency of the corporate and business world, including the existence, ownership and activities of entities, and the people connected with them. OpenCorporates Ltd is registered in England, number 07444723. VAT Number GB 164 3311 33. Registered Office: Aston House, Cornwall Avenue, London N3 1LF, UK. helpdesk@opencorporates.com

Executive summary

- The US is the most important and vibrant market in the world – powered by the activity of millions of companies. Yet the fundamental data about these companies – what companies exist, when they were incorporated, who is connected with them – is very hard to access as data, despite it being public information, collected by the States.

- In today's data-driven world, this dataset is fundamental to almost every aspect of our lives – both business and personal. It is critical for business efficiency, due diligence, innovation, anti-money laundering and corporate transparency, and key to us understanding the complex connected data-driven world we now live in.

- These use cases – and many others – require the dataset to be freely available for all under a licence that allows reuse, i.e. as open data.

- The US States score poorly on access to this data, just 31 out of 100 on average. There are some standout examples – for example Washington state – but there are also many problem states, including Illinois and Delaware. Furthermore, the scores have not increased significantly over the past 3 years, despite significant advances across the rest of the world.

- This has the result of increasing costs for business, undermining competition, and freezing out many groups from easy access to this data, including small and medium sized businesses, civil society, journalists and out-of-state law enforcement.

- It also provides a benign environment for criminals who are now routinely using companies for illicit purposes, including fraud, money laundering and organised crime.

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, from traditional uses such as large corporations and small business, to less savory ones – such as money laundering and organised crime.

Consequently, this dataset – the list of companies which 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 – not just traditional ones such as suppliers, clients, banks, lawyers and accountants, but increasingly with so-called Software as a Service (SaaS) platforms such as Salesforce, Xero, Stripe, GitHub, Mailchimp or Slack – not to mention 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, though social networks, through the ads we see, even through smart speakers and home appliances.

That's why it is so important that everyone – employees, suppliers, customers, government, 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 US context: 50 States + DC

The US is the most important and vibrant market in the world – powered by the activity of millions of companies. Yet astonishingly the fundamental data about these companies – what companies exist, when they were incorporated, who is connected with them, still less what their activities are – is very hard to access as data.

This is largely due to the fact that companies are incorporated by the States (and the District of Columbia), each with its own register, each collecting a different set of non-standard attributes represented in a different way. This generates data silos that massively increase the problems and reduce the utility of the data.

Some states (for example Florida) have made this available as open data for many years. Others, such as Delaware or New Jersey, deliberately restrict access to the data, considering the register's role to be primarily one of revenue generation, as in traditional secrecy jurisdictions such as the Cayman Islands or the British Virgin Islands. This appears to have taken the place of their original role of being a register that records publicly which entities have been incorporated, to provide trust and transparency for a better business environment.

Others put restrictive licences on access to the data – making it unusable for many key use cases.

Together these two issues – siloed non-standardised data and difficulty accessing it as open data – have had significant implications:

- **Black Box data proliferated**
  Black Box data from proprietary vendors proliferated, filling the gap and marginalising the official sources of data. This has increased costs and work, reduced utility and introduced frictions and inefficiencies that distracts business, who are the primary users of company register data, from the real work they should be doing.

- **Official public data is marginalised**
  Official public data from the registers has ceased to be treated as the data of record, and has become secondary to proprietary data, with any identifiers being ignored in favour of proprietary identifiers which are full of data quality problems. A measure of this is that when US state registers do sell data there are very few customers – sometimes just a handful of companies, even for a large state like Illinois. Compare this to the use of UK company register data, for example: UK company numbers are widely used, there are tens of thousands of data users (via free bulk downloads and open API) and it generates benefits to the UK economy of between $1.3 billion and $4 billion a year.

- **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 of apps, SaaS and AI, where innovation and insight comes from combining datasets.

- **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. An example of the second are data-quality issues relating to registrations of out-of-state corporations and LLCs (also known as branches), which are not connected from the company records in the state of incorporation, leading to bad and out-of-date data, as well as lost revenue.
  The result of the consequent data-quality problems is a lack of trust in the underlying data.

- **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 in multiple jurisdictions. The work of law enforcement, tax officers and regulators is made significantly more difficult in the current environment as they are, in practice, restricted to data from their local state.

- **Opaque public procurement**
  Government procurement is made significantly more opaque and inefficient without a master dataset of legal entities.

- **Other stakeholders are shut out**
  As Business Roundtable (an organisation made up of the Chief Executives of the largest corporations in the US) made clear last year, 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.

Foreign Corporations and Foreign LLCs: how it works

A notable feature of the US corporate universe is the widespread use of 'branch' registrations – i.e. registrations in a state by companies incorporated in another state, usually because they have a physical presence or significant business activity in the 'foreign' state.

While branch registrations exist in pretty much all countries, nowhere is it as widespread as in the US (OpenCorporates has matched over 5 million US branches to their home jurisdiction entities). In other countries, it's much more common to form a local subsidiary (i.e. a distinct legal entity) when a company has a significant presence in another jurisdiction.

This use of branches 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.

- Delaware is the primary beneficiary of this situation, with the companies incorporating there having hundreds of thousands of branch registrations in other states.

- 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 States, even though it has significant potential in the areas of tax collection and regulation.

White Box vs Black Box data: why it's important

The canonical source of core company information in the US – which companies exist and existed and what their key attributes are – are the States, yet only a handful of companies consume this data directly. Most use it via intermediaries, 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 new 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.

Identifiers

Identifying companies can be a really difficult problem – and one that has implications for both the utility of the data, and users' trust in it. Company names are highly problematic as identifiers: 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) and can present all sorts of issues in how they are represented (MICROSOFT CORP vs Microsoft Corporation, etc).

As a result, it has become increasingly recognised that identifying companies is critical for good, trusted and useful data. In some countries, the identifier for the company is enshrined in law and must be used in all correspondence. The Global Legal Entity Identifier System 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. Critical to all of these is the idea that identifiers are permanently associated with a single legal entity, contain no information about attributes (which might change, and with it the identifier) and are open and non-proprietary.

In the US, company identifiers are a significant problem, having been rarely given the care and thought that they deserve. In the 9 years that OpenCorporates has been collecting US company data, we have seen over 10 registers change identifiers without any notification and

without any easy or explicit migration from one to another, requiring OpenCorporates to perform complex mapping exercises.

Even more problematic is the widespread use of Dun & Bradstreet's (D&B) proprietary DUNS number. The problems with this number are well-known, including: lack of clarity in definition; the fact that there is no 1:1 relationship between it and legal entities; the proprietary licence; and finally that it locks users into using D&B as a sole supplier of data.

These problems existed for many years, but are particularly problematic in today's data-driven world. As a consequence, the Federal Government last year decided to move away from the DUNS number in procurement.

The results – access to official company data in the US scored

Overview

This report includes the results of our systematic survey of access to US official company register data. Overall, the average score for the US is 32 out of 100. This is very low, equating to registers publishing basic data on the web, including directors, but no open data.

In fact the situation is a little more complex than this, with effectively three categories:

- **Category 1: those that publish only basic data on the web, but publish no open data**
  This is the majority of states, although we have seen an increase in the past 5 years of those publishing open data, as states have realised the benefits of doing so. Often the situation is worse than this as key information – for example director data – may not be in an easily accessible form, but instead in hand-written scanned forms. Alternatively they may publish only the names of agents, but not directors. Both of these situations are problematic, as knowing the prior company history of the directors (or members in the case of LLCs) is critical in many due diligence processes. If you can't see who the individuals who run the company are, and what companies they have been previously involved with, you cannot take an informed view of a company, and whether you want to do business with it.

- **Category 2: those that publish some open data**
  This category possibly only publishes active companies, or leaves out critical information such as the names of officers and directors, or perhaps they don't use an explicit open licence. Historic information is critical when doing investigations, due diligence, or combining the dataset with other data (for example business licences, or

sanctions data).

- **Category 3: those that publish all their data as open data**
  At the moment this is limited to Washington state, who not only provide a bulk data download but also an open API (application programming interface), meaning you can not only easily get the data in bulk, allowing it to be easily analysed and combined with other datasets. Florida and Vermont do publish the data in full but do not use an explicit open licence.

How does this compare with other countries?

The US's score of 31/100 is low, but it is by no means the lowest country or area. A number of countries score 0 points. These are principally either developing countries that have no online register, or secrecy jurisdictions such as the Cayman Islands or British Virgin Islands.

European countries on average get 41.5/100, considerably higher than the US, and with quite a different profile too. In Europe (58 countries), there are 19 that get over 50 and two that get 90/100 (the UK and Denmark). In the US, out of 51 jurisdictions, only 7 get over 50/100 and the top score is 80/100 (Washington state).

In fact, the gap will be soon much bigger, as the European Union has committed to publishing company register data, including company ownership, as open data from 2020. This will increase the score for Europe considerably, as there are a number of countries that score 0/100, as they currently don't even allow free access to the data on the web.

While the US's score is considerably better than Canada's (19/100), this figure is significantly reduced due to the lack of online registers for some of the provinces (which automatically score 0/100).

The US's score is slightly higher than China, which scores 29 out of 100, but less than Russia, which scores 40 out of 100.

The best (and the worst)

As previously noted, Washington state is the best state in the US for access to official company register data. It achieves this position by scoring top marks in the following categories: making data freely available, publishing data freely, using an explicit open licence and making director information available as open data. It also makes the information available both as bulk downloads and via a free and open API – see *Bulk Data vs API* below – which currently doesn't score additional points, but marks the state out as a clear leader in the US.

Below Washington state, a number of states get honorable mentions:

Rank	State	Score (out of 100)
2	Iowa	75
3	Colorado	70
3	Texas	70
5	Alaska	65
6	Oregon	60
7	Virginia	55
7	Vermont	55
9	Wyoming	50

An interesting factor in the US scores is that there is no clear pattern. The West (US Census region) is significantly higher, but this is skewed by high scores for Washington, Oregon, Alaska and Colorado (California scores only 25).

US Region	Average Score
Midwest	27.5
North-East	32.8
South	29.7
West	42.9

Nor can it be seen as a split along any of the following lines: rural vs industrial, big city vs small town, or even stereotypical party political views. In fact, states with Republican Secretaries of State on average score slightly higher than ones with Democrat ones.

A tale of two states: Iowa vs Illinois

An example of this can be seen in two states bordering each other: Iowa and Illinois. Iowa scores a creditable 75 out of 100, making the company register available as an open dataset for anyone to download from their open data portal. It also licences the data with an explicit open licence (requiring just attribution), giving clarity and utility to end users. The Secretary of State is Republican.

Iowa currently only makes the agent details available in the open dataset (not the directors, losing 5 points), and only the active companies are in the open dataset (although if you have the inactive companies, like OpenCorporates does, you can infer when active companies become inactive, and when inactive companies are revived).

Illinois, which has a Democratic Secretary of State, on the other hand, does not publish any open data on companies. Not only that, it explicitly appears to want to actively resist people using the data, with a highly visible notice threatening anyone who tries to use a computer program to collect the data with criminal prosecution. Illinois is the only state whose Secretary of State publishes a threat of this kind. There are also further restrictions on its use in the Illinois Administrative Code.

Finally, while Illinois does sell the data, it's not clear what terms and conditions it is under, as the office of the Secretary of State was not willing to share any information about the terms of use (many of the states that do charge for their data, such as California or Wisconsin, charge only an administration fee for this and don't add restrictions on use).

For all these reasons, we consider Illinois to be the least transparent jurisdiction in the nation for company registration.[1]

Bulk data vs API: which is better?

If the purpose of the 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.

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

[1] Delaware scores five points lower than Illinois in the survey on which this report is based; however, the survey was not originally designed to capture the consequences of threatening criminal prosecution against users. Illinois' legal threat makes its company registration data even less transparent than Delaware's.

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 jurisdictions[2].

How things have changed in the past 3 years

We last performed a systematic survey of access to US official company register data over 3 years ago, and the average score was actually almost identical. While there have been a few improvements (mainly due to an increase in registers publishing open data), there have also been a few decreases too, primarily through identifying data quality issues, or other problems with the data being published.

This lack of improvement is all the more surprising given the increased adoption of open data, both specifically relating to company register data, and to government data in general. For example in that same period, countries as diverse as France, Singapore and Russia have started publishing open company register data, and in the US last year, President Trump signed into law the OPEN Government Data Act, which mandated that "Government data assets made available by an agency shall be published as machine-readable data…in an open format, and…under open licenses."

A key part of the move towards open data in Federal government, both in the current administration and the previous one, has been the improvements in efficiency it allows. By making government data freely available in a standardised form under an open licence, all parties – including government itself – can improve efficiency, quality, and reduce costs, as well as bringing multiple other benefits, such as innovations and services built on top of the data.

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 active[3], whether a company is being impersonated, or even whether it has ever existed at all.

[2] 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.

[3] For example, see: https://www.eveningexpress.co.uk/news/business/firm-at-centre-of-lekoil-scandal-struck-off/

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, 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 the process – 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 friction to business (for example the margins charged by the aggregators, the restrictive licences).

Economically it makes 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.

Benefits to US States of open company data

Due to the distributed nature of company data in the US, and the resulting siloisation of it, there are very significant inefficiencies, costs and other problems that affect all users of company data.

Not least among these users are the States themselves, and they would also be among the biggest beneficiaries of open company data. 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 the States

- **Improved data quality**
  By increasing the number of users and use cases, the registers would see significant improvements in data quality, due to better feedback loops. OpenCorporates itself receives countless number 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 States themselves 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 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 registers as open data addresses all these problems, reducing costs, allowing new suppliers into the market (or allowing the States 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 States are missing out on tax revenue. We understand this is particularly a problem in relation to local entities doing business in other states, and out-of-state entities doing business in state.

- **Better law enforcement and regulation**
  We have heard a number of stories about law enforcement using company register data in innovative ways, including to recover money from parents who are not paying

their child support, and to identify fraudsters.

- **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, fraud and organised crime. Yet this is precisely what they are doing when they allow incorporation or registration with little or no information, and then 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**
  Procurement data in the US has traditionally been oriented around Dun & Bradstreet's proprietary DUNS number, which was first introduced in 1963. However the needs of 2020 are very different from those of the 1960s. The DUNS number's opacity, its opaque data model and the lock-in it creates make it, in many peoples' view, no longer fit for use in procurement (which is fundamentally about contracts between legal entities). A significant step towards this occurred last year, when the Federal Government's General Services Administration announced that it was phasing out the DUNS number in procurement.

- **Enhanced position as the primary source for the data**
  Even where a State register is the sole source of the information sold by legacy proprietary vendors this is not acknowledged due to the Black Box business. As company register data is used in more and more contexts via SaaS applications, the official register is increasingly marginalised, and its reputation and that of the State suffers as a result.

Indirect benefits to the 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 state (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 where you could easily access the data, or one where it was hidden away?

- **More innovation**
  The tiny number of users of US 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 US 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 State that suffers from the illicit use of companies. Business does too.

OpenCorporates will be running pilot programs later this year to help state 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

State	US Census Region	Total (Max 100 points)	Basic data online	Licence	Structured data freely available	Directors	Annual accounts	Share-holders
Alabama	South	**25**	20	5	0	0	0	0
Alaska	West	**65**	20	5	20	10	0	10
Arizona	West	**35**	20	5	0	5	0	5
Arkansas	South	**25**	20	0	0	5	0	0
California	West	**25**	20	5	0	0	0	0
Colorado	West	**70**	20	30	20	0	0	0
Connecticut	North-East	**45**	20	5	20	0	0	0
Delaware	South	**15**	10	5	0	0	0	0
District of Columbia	South	**25**	20	0	0	5	0	0
Florida	South	**45**	20	0	20	5	0	0
Georgia	South	**25**	20	5	0	0	0	0
Hawaii	West	**25**	20	0	0	5	0	0
Idaho	West	**25**	20	5	0	0	0	0
Illinois	Midwest	**20**	20	0	0	0	0	0
Indiana	Midwest	**30**	20	5	0	5	0	0
Iowa	Midwest	**75**	20	30	20	5	0	0
Kansas	Midwest	**25**	20	0	0	5	0	0
Kentucky	South	**25**	20	0	0	5	0	0
Louisiana	South	**30**	20	5	0	5	0	0
Maine	North-East	**25**	20	5	0	0	0	0
Maryland	South	**20**	20	0	0	0	0	0
Massachusetts	North-East	**25**	20	0	0	5	0	0
Michigan	Midwest	**20**	20	0	0	0	0	0
Minnesota	Midwest	**10**	10	0	0	0	0	0
Mississippi	South	**25**	20	0	0	5	0	0
Missouri	Midwest	**30**	20	5	0	5	0	0
Montana	West	**25**	20	5	0	0	0	0
Nebraska	Midwest	**25**	20	0	0	5	0	0
Nevada	West	**25**	20	0	0	5	0	0
New Hampshire	North-East	**25**	20	0	0	5	0	0
New Jersey	North-East	**25**	20	5	0	0	0	0

New Mexico	West	**30**	20	5	0	5	0	0
New York	North-East	**45**	20	5	20	0	0	0
North Carolina	South	**25**	20	5	0	0	0	0
North Dakota	Midwest	**20**	20	0	0	0	0	0
Ohio	Midwest	**25**	20	5	0	0	0	0
Oklahoma	South	**25**	20	5	0	0	0	0
Oregon	West	**60**	20	30	10	0	0	0
Pennsylvania	North-East	**25**	20	0	0	5	0	0
Rhode Island	North-East	**30**	20	5	0	5	0	0
South Carolina	South	**20**	20	0	0	0	0	0
South Dakota	Midwest	**25**	20	5	0	0	0	0
Tennessee	South	**25**	20	5	0	0	0	0
Texas	South	**70**	20	30	20	0	0	0
Utah	South	**25**	20	5	0	0	0	0
Vermont	North-East	**50**	20	5	20	5	0	0
Virginia	South	**55**	20	5	20	10	0	0
Washington	West	**80**	20	30	20	10	0	0
West Virginia	South	**30**	20	5	0	5	0	0
Wisconsin	Midwest	**25**	20	5	0	0	0	0
Wyoming	West	**50**	20	0	20	10	0	0

Methodology

The research for this report was all conducted in November and December 2019, 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.

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 Brazil), 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
- 5 points for no explicit licence
- 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

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 180 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. It has a unique public-benefit business model, with clients and partners including Mastercard, LexisNexis, FactSet and the US Federal Government.

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.