

Digital Agenda Assembly 2012

Brussels, 21-22 June

The closed world of company data

And why we need to open it up



We live in a corporate world

...but much of it is invisible to us
(citizens, journalists, NGOs, regulators,
governments, SMEs, etc...)

Companies no longer look like this



nor even like this

But like this

opencorporates The Open Database Of The Corporate World alph

Search (31 million companies, 35 jurisdictions)

Goldman Sachs Search

SUPPORT OPENCORPORATES Spread the word!

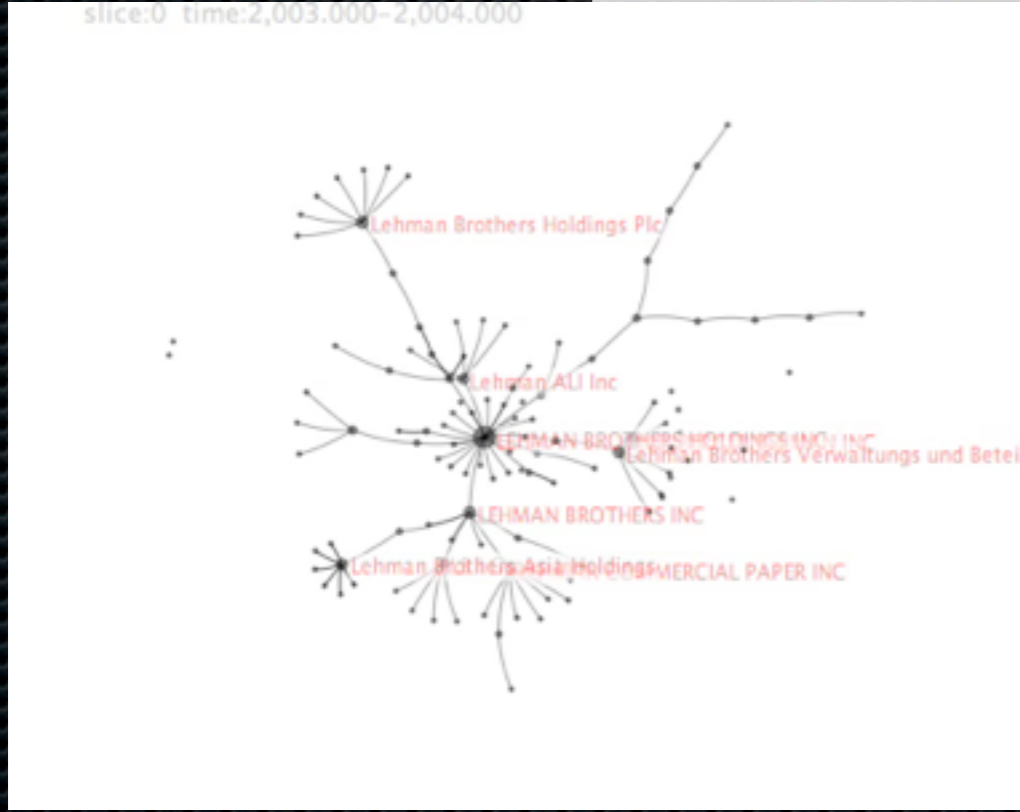
Tweet 0 Like +1 0 +

Found 348 results

- inactive** AMERICA GOLDMAN SACHS INTERNATIONAL GROUP INC. (California (US), 20 Nov 2007-)
- inactive branch** GOLDMAN SACHS & PARTNERS AUSTRALIA CAPITAL MARKETS LIMITED (New Zealand, 27 Oct 1999-11 Mar 2011)
- GOLDMAN SACHS & PARTNERS NEW ZEALAND EQUITY FINANCE LIMITED** (New Zealand, 13 Mar 2002-)
- GOLDMAN SACHS & PARTNERS NEW ZEALAND HOLDINGS LIMITED** (New Zealand, 23 Dec 1988-)
- GOLDMAN SACHS & PARTNERS NEW ZEALAND LIMITED** (New Zealand, 16 Dec 1988-)
- GOLDMAN SACHS & PARTNERS NEW ZEALAND MANAGEMENT LIMITED** (New Zealand, 13 May 1988-)
- PTNERS NEW ZEALAND MEZZANINE LIMITED**
- TNERS NEW ZEALAND PRIVATE EQUITY**

Filter by jurisdiction

- 1 Alaska (US)
- 5 Bermuda
- 11 California (US)
- 1 Canada
- 5 Connecticut (US)
- 19 District of Columbia (US)
- 14 Georgia (US)
- 10 India
- 40 Ireland
- 2 Isle of Man
- 2 Jersey
- 15 Luxembourg
- 4 Maine (US)



or like this

And they are more complex by the day

- ✦ Growing in scale – not 10s of legal entities but 1000s
- ✦ Growing in speed – we are seeing the beginnings of high-frequency company formation
- ✦ Growing in opacity – use of secrecy jurisdictions and off-register entities to provide firewalls to tax, regulation, information
- ✦ Growing in complexity – not a hierarchy but a complex, sometimes even circular network of entities

And this matters

No understanding = no control

- ✦ Leads to systematic problems – Lehmans, pollution exporting, market failures, etc
- ✦ Reduces accountability and corporate governance. Enables and encourages companies to behave like bad corporate citizens
- ✦ Enables of money laundering, organised crime and corruption (see World Bank Puppet Masters report)
- ✦ Remember, companies are artificial entities given legal personality by the state for the good of society

Getting the data matters

- ✦ In the 21st century, data is power
- ✦ We've always been governed by data, now our lives are data
- ✦ Huge asymmetry of access to public data
- ✦ Sold and collected to enrich global proprietary databases, denied to citizens

Int **DATA** ration is the currency
of democracy

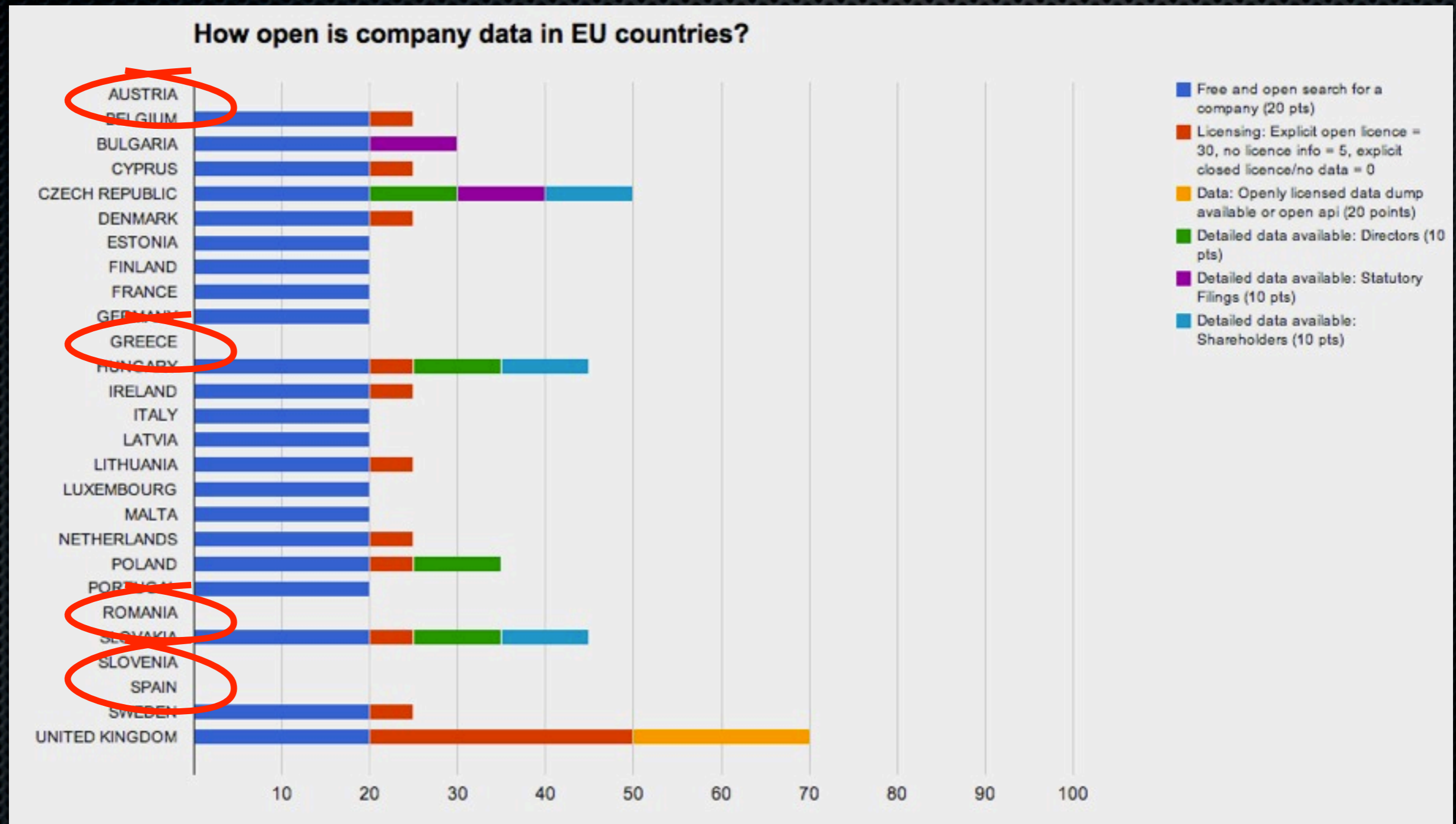
Thomas Jefferson

So how do EU countries score for
to **FAIL!** data?

4 key measures

- ✦ **Basic search:** can you search the company register freely, without charge and without registration
- ✦ **Licence:** Is there a licence that allows open reuse of the information
- ✦ **Data:** Is the information available as open data as a data dump or an API
- ✦ **Depth:** Is there sufficient information to get a true picture of the company and those who control it – directors, significant shareholdings, and statutory filings

The results were not good...



We set out to change that

A simple (huge) goal: build an openly licensed database with an entry (and URI) for every corporate legal entity in the world

The screenshot shows the OpenCorporates website interface. At the top, the logo 'opencorporates' is followed by the tagline 'The Open Database Of The Corporate World' and a search bar. Below the search bar, there are social media sharing options for Twitter (571), Facebook (121), and a '+1' button (31). A prominent statistic states: 'We have information on 35,401,954 companies'. To the right, there is a section for 'Just released: OpenCorporates API' and another for 'Use Google Refine to match companies in your data'. A large red banner is overlaid on the bottom half of the screenshot, containing the text: 'Now over 40 million companies in 52 jurisdictions Including 22 US states'. The background of the screenshot shows a list of jurisdictions with their respective company counts, including Alaska (US), Albania, Aruba, Bermuda, California (US), Canada, Connecticut, Isle of Man, and Jamaica.

Jurisdiction	Count
Alaska (US)	100,189
Albania	37,533
Aruba	43,459
Bermuda	42,193
California (US)	2,547,699
Canada	608,044
Connecticut	656,481
Isle of Man	137,282
Jamaica	82,491

BEMROSE GROUP LIMITED

SUPPORT OPENCORPORATES Spread the word

Tweet 0 Like +1 0

Company Number 03978230

Status Liquidation

Incorporation Date 20 April 2000 (about 12 years ago)

Company Type Private Limited Company

Jurisdiction United Kingdom

Previous Names INHOCH 2059 LIMITED (2000-07-12)

Corporate Grouping

none known. [ADD NOW?](#)

Similarly named companies

- **4IMPRINT GROUP PLC** (United Kingdom, 23 Nov 1921-)
- **4 IMPRINT USA LIMITED** (United Kingdom, 13 Jul 1994-)

Recent filings for BEMROSE GROUP LIMITED

- 2 Mar 2011 **FILING DATED 2 MARCH 2011**
EXTRAORDINARY RESOLUTION TO WIND UP:LIQ. CASE NO.2
- 2 Mar 2011 **FILING DATED 2 MARCH 2011**
STATEMENT OF AFFAIRS/4.19:LIQ. CASE NO.2
- 2 Mar 2011 **NOTICE OF APPOINTMENT OF LIQUIDATOR VOLUNTARY WINDING UP (MEMBERS OR CREDITORS)**
NOTICE OF APPOINTMENT OF LIQUIDATOR (VOLUNTARY):LIQ. CASE NO.2:IP NO.00008023

[see all filings](#)

Source UK Companies House, <http://xml.companieshouse.gov.uk/>, 28 Mar 2012 (UK Crown Copyright)

Add data about this company (website, address, etc)

UK Gazette Notices

APPOINTMENT OF ADMINISTRATORS NOTICE
Appointment of Administrators, *London Gazette*, 29 October 2010

MEETINGS OF CREDITORS NOTICE

Official Register Entries

SEC EDGAR ENTRY
register id: 1297953

5 core uses

1. An open identifying system

- ✦ URIs can be used as common identifiers among a variety of organisations
- ✦ Can be used without reference to OpenCorporates
- ✦ Because they map to the id issued by the company register the corresponding entry in the registry (and associated info) can be found, and vice versa
- ✦ Fits the new W3c/EU Business Vocabulary
- ✦ Can even be used for companies in jurisdiction we haven't yet imported

2. The simple search

- ✦ Not to be underestimated
- ✦ Massively reduces friction (how long will it take you to find and search multiple jurisdictions)
- ✦ Allows what if questions
- ✦ Potentially generates stories in its own right

The screenshot shows the opencorporates website interface. At the top, the logo 'opencorporates' is followed by the tagline 'The Open Database Of The Corporate World' and a search bar containing 'Bain capital'. Below the search bar, it says 'Found 45 results'. A list of results is shown, including 'BAIN CAPITAL (CCD) HOLDINGS (Delaware (US))', 'BAIN CAPITAL (CCD) HOLDINGS MP, LLC (Delaware (US))', and 'BAIN CAPITAL (IRELAND) INTEGRAL INVESTORS LIMITED (Ireland)'. There is a 'Filter by jurisdiction' dropdown menu currently set to 'California (US)'. Below the search results, there is a tweet from Lisa Evans (@objectgroup) that reads: 'Using <http://opencorporates.com> to look up where companies are likely to be paying tax is pretty...'

Store danske virksomheder glade for skattely

Knap hver femte store danske virksomhed har datterselskaber i skattelylande. Mens især rederier og banker benytter sig af selskaber under varmere himmelstrøg som Bermuda og Gibraltar, er virksomheder inden for industri og landbrug fraværende. Samtidig ligger SKAT's udvekslingsaftaler med selvsamme skattelylande næsten ubrugte hen. Opposition kræver flere penge til SKAT.

SELSKABER

Destinationer som Bermuda og Gibraltar er ikke bare attraktive for solhøngrende turister, mens også for veletablerede, danske virksomheder.

Således har næsten hver femte af Danmarks største virksomheder etab-

Ifølge professor ved Aarhus Universitet og ekspert i selskabsbeskatning og skatteret Jan Pedersen dokumenterer oplysningerne, at danske virksomheder er med i den globale kamp om skattefordele. Han understreger, at han ikke kender de konkrete omstændigheder, men siger:

»Man kan jo klart have en formod-

blot er ét af de 130 lande i verden, hvor Mærsk har aktiviteter. Han understreger, at skatteforhold bør være et nationalt anliggende for det enkelte land.

»Selskaberne er ikke oprettet af skattemæssige årsager. A.P. Møller Gruppen driver international virksomhed, der kræver tilstedeværelse i stort set samtlige lande. Derfor ejer gruppen også sel-

3. Source for additional info

- ✦ Addresses, filings, status, websites...
- ✦ Intl trademarks, UK govt spending, official notices, health & safety violations...
- ✦ Other IDs: SEC, CAGE, etc – allows reverse mapping queries, e.g. show me legal entity mapped to a CLK code

opencorporates The Open Data

EXPERIAN LIMITED

Company Number 00653331

Status Active

Incorporation Date 22 March 1960 (almost 52 years ago)

Company Type Private Limited Company

£1,163 from [Rutland County Council](#) on 2011-03-14 ([details](#)) Mark text:

£900 from [Brighton and Hove City Council](#) on 2011-07-15 ([details](#)) **INTERNATIONAL** Mark text:

£20 from [Brighton and Hove City Council](#) on 2011-07-29 ([details](#)) **INTERNATIONAL** Mark text:

£11 from [Brighton and Hove City Council](#) on 2011-07-29 ([details](#)) **INTERNATIONAL** Mark text:

Company Addresses

COMPANY ADDRESS
EXPERIAN, LANDMARK HOUSE, EXPERIAN WAY, NG2
BUSINESS PARK, NOTTINGHAM, NG80 1ZZ, United Kingdom

TRADING ADDRESS
Cardinal Place, 80 Victoria Street, London, SW1E 5JL

Approved
APPROVED

Official
UK DATA
register id:

4. Reconciliation (matching names to legal

Clean up messy
company names
(& prev names)
to legal entity,
and from there
to other data

Google Refine
reconciliation
service (specific
to jurisdiction)

The screenshot shows the Google Refine reconciliation interface. On the left, the 'Facet / Filter' panel is active for 'Company', showing 787 records. The 'Donor type' facet is selected, with options: 'Company' (787), 'Friendly Society' (219), 'Limited Liability Partnership' (146), and '(blank)' (1142). The 'Company reg. no.' facet is also visible with 1 choice: 'true' (1566). The main table displays 787 matching records out of a total of 33821. The table has columns for 'All', 'EC reference', 'Entity name', 'Entity type', and 'Regulated donee type'. The records are sorted by name count. The first few records are:

All	EC reference	Entity name	Entity type	Regulated donee type
★	2253. C0002254	The Rt Hon Peter Hain MP	Regulated Donee	Leadership Candidate
★	2410. C0002411	Mr Boris Johnson MP	Regulated Donee	MP - Member of Parliament
★	3046. C0003046	Mr Oliver Letwin	Regulated Donee	MP - Member of Parliament
★	3056. C0003058	Ms Philippa Stroud	Regulated Donee	Candidate
★	3064. C0003066	Ms Philippa Stroud	Regulated Donee	Candidate
★	3114. C0003116	Mr Douglas Alexander MP	Regulated Donee	MP - Member of Parliament
★	3183. C0003185	The Rt Hon	Regulated	Leadership Candidate

5. The platform

- ✦ API: allows all information to be retrieved as data, even searches
- ✦ Users can now add data too
- ✦ Coming soon: the option to match data to companies

The screenshot displays the OpenCorporates API website. At the top, the logo 'opencorporates' is followed by 'The Open Database Of The Corporate World' and 'API'. A 'REST API introduction ::' header is visible, along with a 'Contributor: James' section indicating a join date of November 24, 2011. Below this, there's a search bar and a navigation menu. The main content area features a search result for 'MAERSK corporate grouping *', listing 205 companies. A sidebar on the right provides 'Latest data for these companies', including a 'UK GAZETTE NOTICE' and a 'COMPANY ADDRESS' in Middlesbrough. At the bottom, a 'Vat Number' search result is shown for 'International Training Service (ITS) Ltd' with the number 238840151. Social media sharing options like 'Like' and 'Tweet' are present throughout the page.

How have we done it?

- ✦ Co-operation – we get data direct from some company registers (UK, NZ, a few US), and are working with international institutions (EC, W3c, Financial Stability Board, etc) to improve visibility and reuse of company info
- ✦ Community – a lot of the data has been contributed by the open data community (thanks, ScraperWiki)
- ✦ Cool open-source software (100% open source platform/tools)
- ✦ Colossal scraping (100,000s of pages/API calls per day)

What next?

- ✦ Recently started adding company directors and officers
- ✦ More public data – political donations, lobbyists, other ID systems
- ✦ Relationships between corporate entities
- ✦ More options for community to add/curate data