

OpenCorporates Data Dictionary

Extended Companies & Officers

VERSION HISTORY

Version	Date	Prepared by	Comments
1.0	14-Sep-2016	Alex Skene	Initial version
1.1	06-Dec-2016	Alex Skene	Companies: - Columns re-ordered & home_jurisdiction_text included to meet revised common Exporter spec [OCD-267]
1.2	09-Sep-2017	Alex Skene	Companies: - Increased field length of: name, normalised_name, company_type, current_status Officers: - Increased field length of: name, position

OpenCorporates Data Dictionary - Extended company data

Description

This data is core company data from OpenCorporates master company database. Given that companies have 0..n officers they are in a separate dump, and have a separate data dictionary.

All fields are optional unless shown as Mandatory

Primary Key

company_number, jurisdiction_code

field name	format	field description	Field-level notes
company_number	String - varchar(128)	Identifier given by the Company register to the entity. Together with the jurisdiction_code this forms a composite primary key for the companies dataset.	Mandatory
jurisdiction_code	String - varchar(8)	Code for the jurisdiction (lowercase, underscore version of ISO 3166-2 code). List of jurisdiction codes and the names of the jurisdictions is available at https://api.opencorporates.com/jurisdictions	Mandatory
name	String - varchar(2000)	Legal name of entity	Mandatory
normalised_name	String - varchar(2000)	Normalised legal name of entity, using OpenCorporates normalisation.	
company_type	String - varchar(1000)	Type of entity (as described by the company register)	
nonprofit	Boolean	Whether the company_type indicates it is a nonprofit	OpenCorporates mapping based on company_type
current_status	String - varchar(1000)	Status of entity at retrieved_at date as reported by company registry	
incorporation_date	Date (ISO 8601: YYYY-MM-DD)	Date of incorporation	
dissolution_date	Date (ISO 8601: YYYY-MM-DD)	Date of dissolution of entity	
branch	Enum - 'F', 'L' or null	If this is a local registration of a foreign company (usually called a branch or foreign corporation) this will be 'F'. In a small number of cases we also have some local branches of entities, and these have an 'L' flag	
business_number	String Size limit - see Addtl Info sheet - Note [1]	Cross-government business numbers for those jurisdictions (e.g. Australia, New Zealand, Canada) that have them	
current_alternative_legal_name	String Size limit - see Addtl Info sheet - Note [1]	Alternative legal name (in another language)	
current_alternative_legal_name_language	String Size limit - see Addtl Info sheet - Note [1]	Language of the alternative name, if known	
home_jurisdiction_text	String Size limit - see Addtl Info sheet - Note [1]	Jurisdiction of the home company, in cases where the entity is a branch registered in a different jurisdiction to the company, else will be the jurisdiction of the company matching the company file's jurisdiction	
native_company_number	String Size limit - see Addtl Info sheet - Note [1]	On occasion a company number is re-used or renumbered by a registry. This field stores the former number	

previous_names	String: List of names separated by Pipe character (' ') Size limit - see Addtl Info sheet - Note [1]	Previous names for the entity	
alternative_names	String: List of names separated by Pipe character (' ') Size limit - see Addtl Info sheet - Note [1]	Alternative names for the entity	
retrieved_at	Datetime (ISO 8601)	Datetime the record was retrieved from the official source	
registry_url	String - text	URL of the page for the company on the official corporate register	
restricted_for_marketing	String - text	Indicates whether marketing is allowed to this customer (as per information provided by the official corporate register)	
inactive	Boolean	Whether the company status indicates the company is in an inactive state. Most of these are companies that are dissolved in some form or other. It excludes 'dormant' or companies in liquidation, as these can both be actively trading	OpenCorporates mapping based on current_status values
accounts_next_due	Date (ISO 8601)	Date accounts are next due	
accounts_reference_date	dd/mm	Accounting reference date	
accounts_last_made_up_date	Date (ISO 8601)	Date accounts were last made up	
annual_return_next_due	Date (ISO 8601)	Date annual return next due	
annual_return_last_made_up_date	Date (ISO 8601)	Date returns were last made up	
has_been_liquidated	Boolean	Whether company has been liquidated	
has_insolvency_history	Boolean	Whether company has insolvency history	
has_charges	Boolean	Whether company has charges	
registered_address.street_address	String - text	Street address part of address object for registered address	
registered_address.locality	String - varchar(255)	Locality (e.g. city) part of address object for registered address	
registered_address.region	String - varchar(255)	Region (e.g. state) part of address object for registered address	
registered_address.postal_code	String - varchar(255)	Postal code (e.g. ZIP) part of address object for registered address	
registered_address.country	String - varchar(255)	Country part of address object for registered address	
registered_address.in_full	String - text	Registered address of the entity (sometimes called service address). This is represented as a single string – in some cases we have this as an address object, in which case this will be a concatenated version of this.	
home_jurisdiction_code	String - varchar(8)	Jurisdiction code of the home company, in cases where the entity is a branch. See description of jurisdiction_code for details	
home_jurisdiction_company_number	String - varchar(128)	Company number of the home company (i.e. the incorporated legal entity) if the entity is a branch. This has been reconciled by OpenCorporates. Not all branches are reconcilable to home companies	

industry_codes	String - text: List of SIC codes separated by Pipe character (' ') See lookup table for actual lengths	Standard Industrial Classifications	These are concatenated codes, including code scheme and code number. We can provide a lookup table for these (See http://api.opencorporates.com/documentation/API-Reference#industry_codes)
latest_accounts_date	Date (ISO 8601)	Date of latest filed accounts	
latest_accounts_cash	Integer Size limit - see Addtl Info sheet - Note [1]	Cash at bank in latest accounts, in local currency	
latest_accounts_assets	Integer Size limit - see Addtl Info sheet - Note [1]	Current assets in latest accounts, in local currency	
latest_accounts_liabilities	Integer Size limit - see Addtl Info sheet - Note [1]	Current liabilities in latest accounts, in local currency	

Data Dictionary - Core officer data

Description

This data is officer data associated with OpenCorporates master company database. We have a distinct record for each position, so that if a person has 3 roles (director, secretary, president) they will have 3 records.
All fields are optional unless shown as Mandatory

Primary Key

id

field name	format	field description	Field-level notes
id	Integer	Internal OpenCorporates ID for officership	Mandatory
company_number	String - varchar(128)	company number	Mandatory. FK to Companies
jurisdiction_code	String - varchar(8)	Code for the jurisdiction (see companies data dictionary). Together with the jurisdiction_code this forms a composite foreign key for the company	Mandatory. FK to Companies
name	String - varchar(2000)	Full name of the officer	Mandatory
title	String Size limit - see Addtl Info sheet - Note [1]	Title of the office, e.g. "Mr"	
first_name	String Size limit - see Addtl Info sheet - Note [1]	First name of officer	
last_name	String Size limit - see Addtl Info sheet - Note [1]	Last name of the officer. In the case of corporate officers, the name of the company will appear in this field	
position	String - varchar(1000)	Position held	
start_date	Date (YYYY-MM-DD)	Start date of officership	
person_number	String Size limit - see Addtl Info sheet - Note [1]	Companies House uid for the person (i.e. individual), including address change counter. e.g. 204942900001 has a counter of "0001". This number is incremented when the address changes. This number is not a unique uid for individuals, as the address counter may be different for the person for the different companies they are associated with	Applies to UK data only
person_uid	String Size limit - see Addtl Info sheet - Note [1]	Derived person uid from the person_number. Individuals with the same person uid (e.g. "20494290") are asserted by Companies House to be the same person	Applies to UK data only
end_date	Date (YYYY-MM-DD)	End date of officership	
current_status	String Size limit - see Addtl Info sheet - Note [1]	Status of the officership as reported by the company register (if given)	
occupation	String Size limit - see Addtl Info sheet - Note [1]	Occupation of director if given	

nationality	String Size limit - see Addtl Info sheet - Note [1]	Nationality of director	
country_of_residence	String Size limit - see Addtl Info sheet - Note [1]	Country of Residence of the officer	
partial_date_of_birth	Partial Date (YYYY-MM)	Partial date of birth, e.g. "1974-04"	
type	Enum	Person or "Company" or null	
address.in_full	String Size limit - see Addtl Info sheet - Note [1]	Contact address of the director	
address.street_address	String Size limit - see Addtl Info sheet - Note [1]	Street address portion of contact address	
address.locality	String Size limit - see Addtl Info sheet - Note [1]	Locality (e.g. City) portion of contact address	
address.region	String Size limit - see Addtl Info sheet - Note [1]	Region (e.g. County) portion of contact address	
address.postal_code	String Size limit - see Addtl Info sheet - Note [1]	Postcode portion of contact address	
address.country	String Size limit - see Addtl Info sheet - Note [1]	Country portion of contact address	
retrieved_at	Datetime (ISO 8601)	Datetime the record was retrieved from the official source	
source_url	String - text	URL of the page for the officer on the official corporate register	

Additional notes

File / Folder naming	- Each delivery will be in a folder named as per "as-at" date - eg "2016-06-01" - Sample / test files will be in separate dated folders - eg "sample_2016_06_01"
Compression	- Files will be compressed in gzip format, with ".gz" extension
".ok" files	- One OK file will be produced for each data file - Each OK file will contain only the number of records for the associated file on a single line, with no headers - eg: 1324 - OK file will be made available on SFTP site when data files have completed copying (to help prevent downloading of partial files)
md5sum	- File named "md5sum" will be added showing a MD5 checksum for all gzipped files - this helps check for any integrity issues during download
File transfer mechanism	- SFTP - User authentication: choice of password or private-key certificate
Format of file	- CSV, conformant to RFC 4180 https://tools.ietf.org/html/rfc4180 - Linebreaks in text indicated with \n (common in address fields) - Unix line endings
Column separator	, (comma)
Fields enclosed by	Optionally enclosed by double quotes, e.g when field contains a comma e.g: xxx,"Test Company, LLC",yyy
Escape character	Double quotes in text are escaped with double quotes as per RFC 4180 e.g: xxx,"test ""hello"" company llc",yyy
Encoding	UTF-8

Data Types

varchar(255)	string up to 255 chars in length
text	string up to 65,535 bytes in length - NB: UTF8 consumes 1-3 bytes per character
mediumtext	string up to 16MB bytes in length - NB: UTF8 consumes 1-3 bytes per character
pipe separated fields	Unable to determine field length. Data in this field can consist of 0, 1 or multiple instances of an underlying field. so data may need to be normalised during ETL
boolean	contains: true, false, or is left blank to signify "unknown" state

Note [1] Data in this field is combined with other data attributes in YAML serialized format, and is stored in a single mediumtext MySQL field. This allows OpenCorporates to handle a wide variety of string lengths from multiple data sources