

# Introduction to Reference Data in Vortexa Python SDK

When working with Vortexa's Python SDK, one of the foundational components you'll encounter is reference data. But what exactly is reference data, and why is it crucial for your analytics and data manipulation tasks?

## What is Reference Data?

Reference data, also known as master data, encompasses the core datasets that define and categorize the entities within Vortexa's ecosystem. These datasets are used to standardize and provide context to the event data you'll be working with. In essence, reference data acts as the backbone for ensuring consistency, accuracy, and reliability in your data operations.

## Three main Reference Data in Vortexa

**1. Geographies:** Detailed information about various locations, such as ports, terminals, and regions. This allows you to accurately track and analyze movements and activities related to these locations.

**2. Vessels:** Comprehensive data about vessels, including their names, types, sizes, and other attributes. This is essential for tracking vessel movements, understanding fleet compositions, and analyzing shipping trends.

**3. Products:** Information about different types of products, their classifications, and hierarchies. This helps in analyzing trade flows, market dynamics, and commodity-specific trends.

## Import Libraries

```
In [1]: import vortexasdk as v
import pandas as pd
```

```
C:\Users\OuiWeinJien\anaconda3\lib\site-packages\pandas\core\arrays\masked.py:60: UserWarning: Pandas requires version '1.3.6' or newer of 'bottleneck' (version '1.3.5' currently installed).
  from pandas.core import (
```

To start with, let's explore what you could extract from Geographies endpoint.

```
In [2]: # To load all geographies
all_geographies = v.Geographies().load_all().to_df()

# To load all geographies with a specific type (e.g. country)
country_list = v.Geographies().search(filter_layer=['country']).to_df()

# To load all geographies with a name containing 'China'
china_list = v.Geographies().search(term='China').to_df()
```

```
2024-09-24 17:48:31,321 vortexasdk.client - WARNING - You are using vortexasdk version 0.73.0, however version 0.74.1 is available.
You should consider upgrading via the 'pip install vortexasdk --upgrade' command.
```

```
In [3]: all_geographies.head(5)
```

Out[3]:

	id	name	layer
0	0eb7b43e3d4e62db74e187bc4eadadfb878a210201e14a...	21st Century Shipbuilding	[terminal]
1	b27430b57d617a855d44f91fb70441bae69c19a3c0deb4...	2x1 Holding Cape Midia Shipyard	[terminal]
2	d38a8f7bf8ed422b439ad5270be65b60b964bed9568936...	A Pobra Do Caraminal [ES]	[port]
3	3ebcf6f2e43e3a8b06e9b0ae31df3d87f3fbc8d032b72d...	A and P Group Falmouth Shipyard	[terminal]
4	0c8a30f40639257e5352e2c6ac52af2f93b2c6bfed7187...	A and P Group Tees Shipyard	[terminal]

In [4]: `country_list.head(5)`

Out[4]:

	id	name	layer
0	1cd3c07221f9e9b3296c859d0bcd3da17ac6072bfdcc84...	Afghanistan	[country]
1	5e4e7b5040b933b5a0f0d2357ed27ebec432c749b9d63a...	Albania	[country]
2	87269b28eaea324d2c35e97b0ecc837ebc9a244faf94e2...	Algeria	[country]
3	f4435d4fffa5b2ba7a340e3a8e7d421f619d9f7832fa0c...	American Samoa	[country]
4	db3cb74043b8fd438087a3e0e04e3e498b78c3c9790fce...	Andorra	[country]

In [5]: `china_list.head(5)`

Out[5]:

	id	name	layer
0	934c47f36c16a58d68ef5e007e62a23f5f036ee3f3d1f5...	China	[country]
1	a63890260e29d859390fd1a23c690181afd4bd152943a0...	North China	[alternative_region]
2	781cacc7033f877caa4b4106d096b74afe006a96391bf5...	South China	[alternative_region]
3	ae0f224030f7337d0ffe5a54d290a9f0bd029f636eaf12...	China Yangfan Group	[terminal]
4	9a021f43c397b175ddfff7a91d46ee6e6e16d37e9f9d52...	China Steel Chemical	[terminal]

This shows how you could extract the ids, which may be required from other endpoints. In addition, to extract more information about the locations such as centroids, hierarchies etc, we can do `.to_df(columns = 'all')` method.

In [6]: `china_list_enhanced = v.Geographies().search(term='China').to_df(columns = 'all')`  
`china_list_enhanced.head(5)`

Out[6]:		id	name	layer	leaf	parent	exclus
0	934c47f36c16a58d68ef5e007e62a23f5f036ee3f3d1f5...		China	[country]	False	[{'name': 'Far East', 'layer': ['alternative_r...	'Chin ['
1	a63890260e29d859390fd1a23c690181afd4bd152943a0...		North China	[alternative_region]	False	[{'name': 'China (excl. HK & Macau)', 'layer':...	'Nort ['al
2	781cacc7033f877caa4b4106d096b74afe006a96391bf5...		South China	[alternative_region]	False	[{'name': 'China (excl. HK & Macau)', 'layer':...	'Sout ['al
3	ae0f224030f7337d0ffe5a54d290a9f0bd029f636eaf12...		China Yangfan Group	[terminal]	True	[{'name': 'Zhoushan [CN]', 'layer': ['port', '...	'China Grou
4	9a021f43c397b175ddfff7a91d46ee6e6e16d37e9f9d52...		China Steel Chemical	[terminal]	True	[{'name': 'Kaohsiung [TW]', 'layer': ['port', '...	'Ch C 'li

Now we have demonstrated how to extract reference data via Geographies. Similarly, the methodology works for Products & Vessels endpoint as well.

```
In [7]: # To load all products
products = v.Products().load_all().to_df()
products.head(5)
```

Out[7]:	id	name	layer.0	parent.0.name
0	887940a6cf2d527a20d82a5f163ecce502878ceb1cd59f...	0.005 / 50ppm	grade	Gasoil
1	8bb096fb847f92af86235002b2a78ca0437543722cdb8c...	0.05 / 500ppm	grade	Gasoil
2	35f8222ff81fe5befafee9c64c1d76618e4cc53e74021a...	0.1 / 1000ppm	grade	Gasoil
3	881be476857ff08dcf6a8708a2fc279d26770cecf245f7...	0.1+ / 1000ppm Plus (HS)	grade	Gasoil
4	a6ef13d2f3145a1b67a81300c1cfa4f21874f24fab4f8f...	180 CST	grade	High Sulphur Fuel Oil

```
In [8]: # To load Gasoil only
gasoil = v.Products().search(term='Gasoil').to_df()
gasoil.head(5)
```

Out[8]:	id	name	layer.0	parent.0.name
0	b2034f1ad3a4ac269e962f00b9914d6b909923cf904d99...	Gasoil	category	Diesel/Gasoil
1	deda35eb9ca56b54e74f0ff370423f9a8c61cf6a3796fc...	Diesel/Gasoil	group_product	Clean Petroleum Products
2	e06296595e1d554008a70172440d5582c923bdb8182af5...	Coker Gasoil	grade	Dirty Feedstocks
3	feb8190865392ab6caecd7709077a58645ec4828c23d94...	Marine Gasoil	grade	Gasoil

```
In [9]: # To load all vessels
vessels = v.Vessels().load_all().to_df()
```

```
vessels.head(5)
```

Out[9]:

	id	name	imo	vessel_class
0	62f3f3c1f5a663d621fe6cf9537c7d936b547497932f5d...	\tATHINEA	9291248.0	oil_lr2
1	e6b259c04da30a57db353665e7e61f67a0a3222b96c457...	\tBORA	9276004.0	oil_mr2
2	f351708121bce4d357ac5fad967cb1bf7fe5072773f05a...	0051-04		oil_coastal
3	1761da4fb069cd6ce153b6ad1c48e15cdb994eb386e4aa...	058		oil_coastal
4	c817b5994efe14621949533d6777b22ce11db1c6bf9e48...	1011		oil_coastal

```
In [10]: # To Load all vessels with a name containing 'Maersk' (currently named or previously named)
maersk_vessels = v.Vessels().search(term='Maersk').to_df()
maersk_vessels.head(5)
```

Out[10]:

	id	name	imo	vessel_class
0	00d89be99f08890c9122c326aa32c83ae6d557629bc124...	VS REMLIN	9252307	oil_mr1
1	03c191caf28a554c8c1adfc11ff2a7a08ad5fa21a83892...	SOUTH LOYALTY	9537769	oil_vlcc
2	07e562b509f2617c18f9e848c51bdec8e97488c4a41bbc...	HENRIETTE MAERSK	9399349	oil_mr1
3	085ab83b592713e314070620a8cb9f4d795a2b486075f4...	VS	9252292	oil_mr1
4	09332aaa6067574eac586030b5b81cf5554337c4f48d17...	BULL SULAWESI	9180920	oil_lr2

## Conclusion

In this tutorial, we covered the essentials of working with reference data using the Vortexa Python SDK. We began by discussing the importance of reference data and its role in supporting accurate and consistent data operations. Through various examples, including locations, vessels, and products, we demonstrated how reference data can be effectively applied in your analyses.

You've learned how to query and retrieve reference data based on different criteria such as name or type. Mastering the use of reference data enables you to drive more accurate insights, improve data consistency, and enhance your understanding of the energy markets.