

Playing with APIs

How to use APIs to collect your
datasets?

Zhixin (Ivy) Zheng

Agenda

01

Preparation before Using API

02

Hands-on Recipes for **Trove** API

03

A Glance on **Factiva** API

04

Tips in Using API

Preparation before Using API

A few things you need to know before using API...



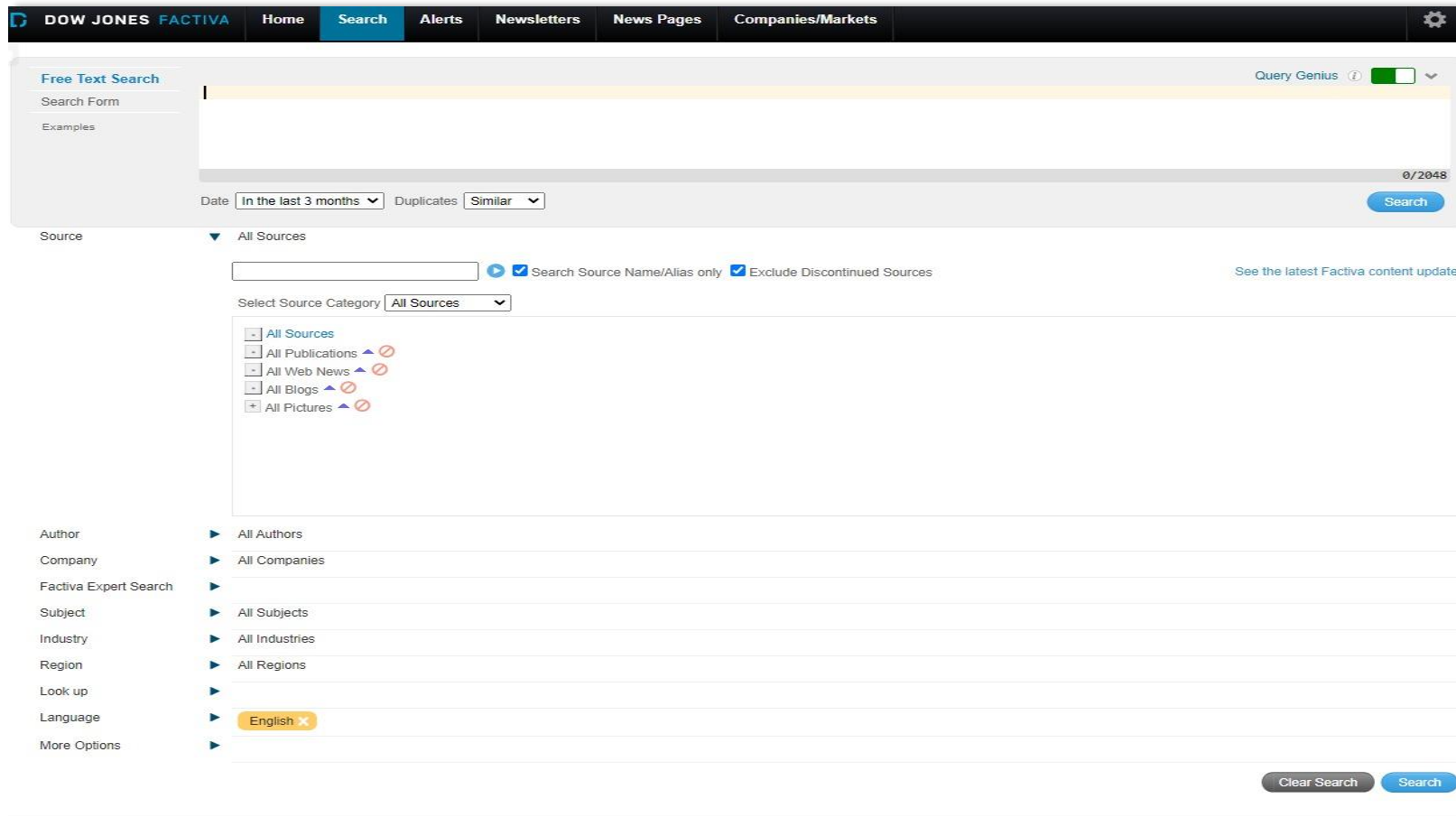
THE UNIVERSITY OF
MELBOURNE

What is API?

API stands for
Application
Programming
Interface.

It is a set of definitions and **protocols** that enables data transmission between one software product and another.

GUI (Graphical User Interface) for Factiva Searching



The screenshot displays the Factiva search interface. At the top, a navigation bar includes links for DOW JONES, Factiva, Home, Search, Alerts, Newsletters, News Pages, and Companies/Markets. The 'Search' tab is active. Below the navigation bar, the 'Free Text Search' section is visible, featuring a search form and a 'Query Genius' indicator. The search criteria are set to 'In the last 3 months' for the date and 'Similar' for duplicates. The source selection is set to 'All Sources'. The search filters are expanded, showing a list of categories: All Sources, All Publications, All Web News, All Blogs, and All Pictures. The search results are displayed in a table with columns for Author, Company, Factiva Expert Search, Subject, Industry, Region, Look up, Language, and More Options. The 'Language' filter is set to 'English'. The search results are currently empty, and the 'Search' button is visible at the bottom right.

DOW JONES Factiva Home Search Alerts Newsletters News Pages Companies/Markets

Free Text Search

Search Form

Examples

Query Genius

0/2048

Date In the last 3 months Duplicates Similar Search

Source All Sources

Search Source Name/Alias only Exclude Discontinued Sources See the latest Factiva content updates

Select Source Category All Sources

- All Sources
- All Publications
- All Web News
- All Blogs
- All Pictures

Author All Authors

Company All Companies

Factiva Expert Search

Subject All Subjects

Industry All Industries

Region All Regions

Look up

Language English

More Options

Clear Search Search

API (Application Programming Interface) for Factiva Searching

```
import json

headers = {"content-type": "application/json", "user-key": "{USER_KEY}"}

request_body = {
    "query": {
        "where": "language_code='en'",
        "includes": {
            "company_codes": [
                "exxn"
            ] # only documents that contain company_code "exxn" will be returned
        },
        "limit": 10000,
    }
}

snapshot_url = "https://api.dowjones.com/alpha/extractions/documents"
requests.post(snapshot_url, data=json.dumps(request_body), headers=headers)
```

Interfaces designed for communicating with program.

GUI

For human beings

✓ **Easier to use!**

- ✓ Capture larger datasets automatically!
- ✓ Export data to your own applications (databases, websites, analysis software, etc)

For computer programs only

API

What Protocol does API follow?

The **H**ypertext **T**ransfer **P**rotocol
(HTTP) works as a request-response
protocol between a client and server.

API follows HTTP, so we can use HTTP
method to send our request to servers.

HTTP Methods:

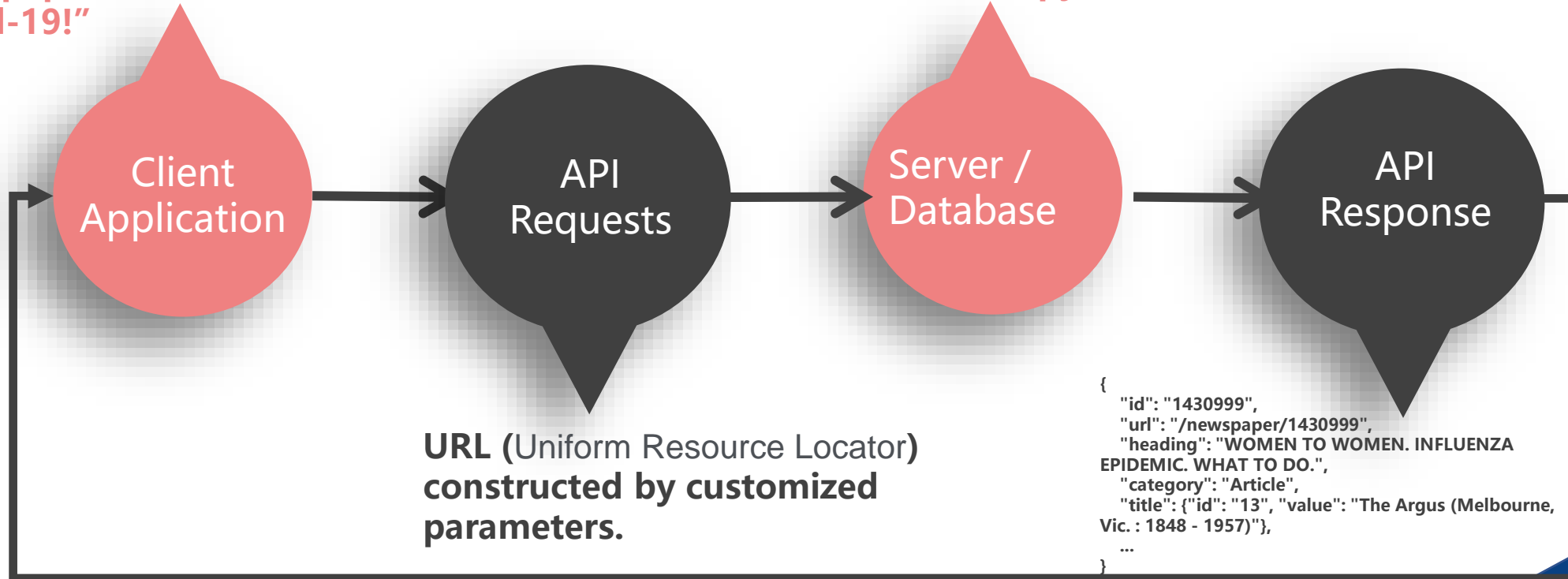
- **GET:** request data from a specified resource
- **POST:** send data to server to create/update
a resource
- PUT, HEAD, DELETE, PATCH, OPTIONS



How does API work?

User: "Hey Trove, I want some newspaper articles about covid-19!"

Server: "Copy that!"



A Simple URL Example from Trove

`https://api.trove.nla.gov.au/v2/result?n=20&q=covid-19&zone=newspaper&encoding=json&key=your-api-key`

Fixed part (points to `https://api.trove.nla.gov.au/v2/result?`)

The number of results to return per zone. (points to `n=20`)

Searching term (points to `&q=covid-19`)

Searching zone (points to `&zone=newspaper`)

Return format (points to `&encoding=json`)

API key (points to `&key=your-api-key`)

Json Example

Attribute

Value

Array data-type

The attribute can have more details

```
{
  "id": "258999523",
  "url": "/newspaper/258999523",
  "heading": "Defence acts on COVID-19 pandemic",
  "category": "Article",
  "array_type_attribute": ['value1', 'value2', 'value3'],
  "title": {"id": "1672", "value": "Air Force News (National : 1997 - 2020)"},
  "date": "2020-03-19",
  "page": 3,
  "pageSequence": 3,
  "relevance": {"score": "1439.3022", "value": "very relevant"},
  "snippet": "IN THE wake of the World Health Organisation declaring Novel Coronavirus (COVID-19) a",
  "troveUrl": "https://trove.nla.gov.au/ndp/del/article/258999523?searchTerm=covid-19",
}
```

	id	url	heading	category
0	258999523	/newspaper/258999523	Defence acts on COVID-19 pandemic	Article

A format using attribute-value pairs and array data-types to store and transport data.

Toolkit for using API

JupyterLab

A python Integrated
Development
Environment

Requests

A Python library helps
you to handle your
request.

Json

A Python library used for
JSON encoding and decoding.

Hands-on Recipes for TROVE API

TROVE

What can you get with Trove API?

Name of Zone	Number of Records	Type of Records	Time Range
List	99,958	List	2010-2021
Gazette	3,535,033	Newspaper & Gazette	1832-2012
Newspaper	233,162,088	Newspaper & Gazette	1803-2020
People	1,291,668	Work	-
Map	335,916	Work	1800-2021
Picture	5,342,126	Work	1820-2021
Collection	4,428,551	Work	1818-2017
Book	17,520,474	Work	1819-2021
Music	2,846,358	Work	1000-2021
(Journal) Article	12,163,390	Work	1800-2021

- 3 types of records across 10 zones;
- Since 19th century;
- Metadata + Full Text (newspaper & gazette only)

03

Look up other
associated data in Trove.

Get information about
a single item in Trove.

02

01

Search across the records in Trove.

**What can Trove
API do for you?**

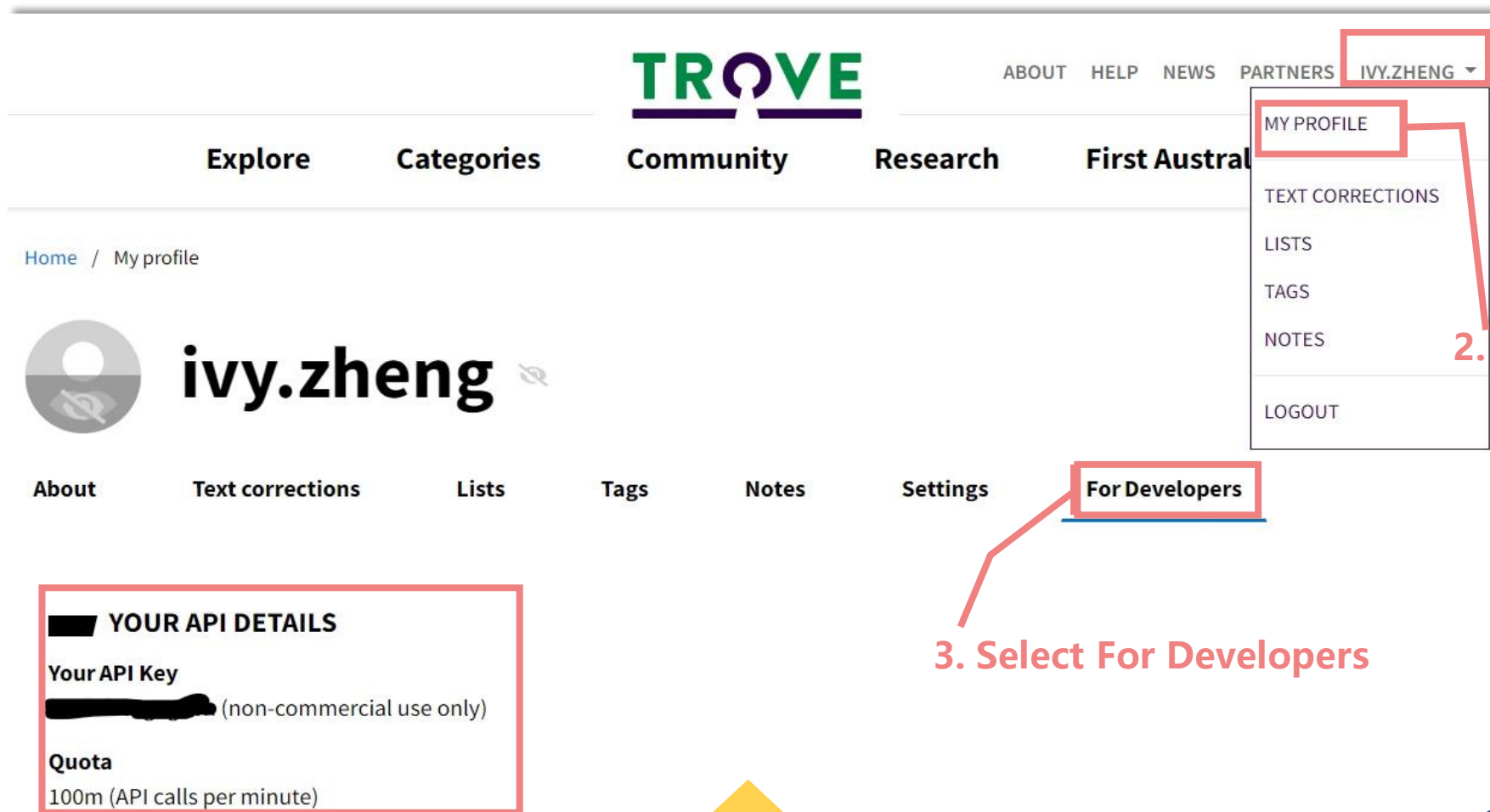
02

Now, it's play time!!

But wait...

You need to apply an API key first!

How to apply a Trove API key?



The screenshot shows the Trove website interface. At the top, the Trove logo is centered, with navigation links: ABOUT, HELP, NEWS, PARTNERS, and a dropdown menu for IVY.ZHENG. Below the logo, there are tabs: Explore, Categories, Community, Research, and First Austral. The main content area shows the user's profile for ivy.zheng. A dropdown menu is open, showing options: MY PROFILE, TEXT CORRECTIONS, LISTS, TAGS, NOTES, and LOGOUT. The 'For Developers' link is highlighted in the bottom navigation bar. A red box highlights the 'YOUR API DETAILS' section, which contains the API key and quota information.

Trove

ABOUT HELP NEWS PARTNERS IVY.ZHENG

Explore Categories Community Research First Austral

Home / My profile

ivy.zheng

About Text corrections Lists Tags Notes Settings **For Developers**

YOUR API DETAILS

Your API Key
[Redacted] (non-commercial use only)

Quota
100m (API calls per minute)

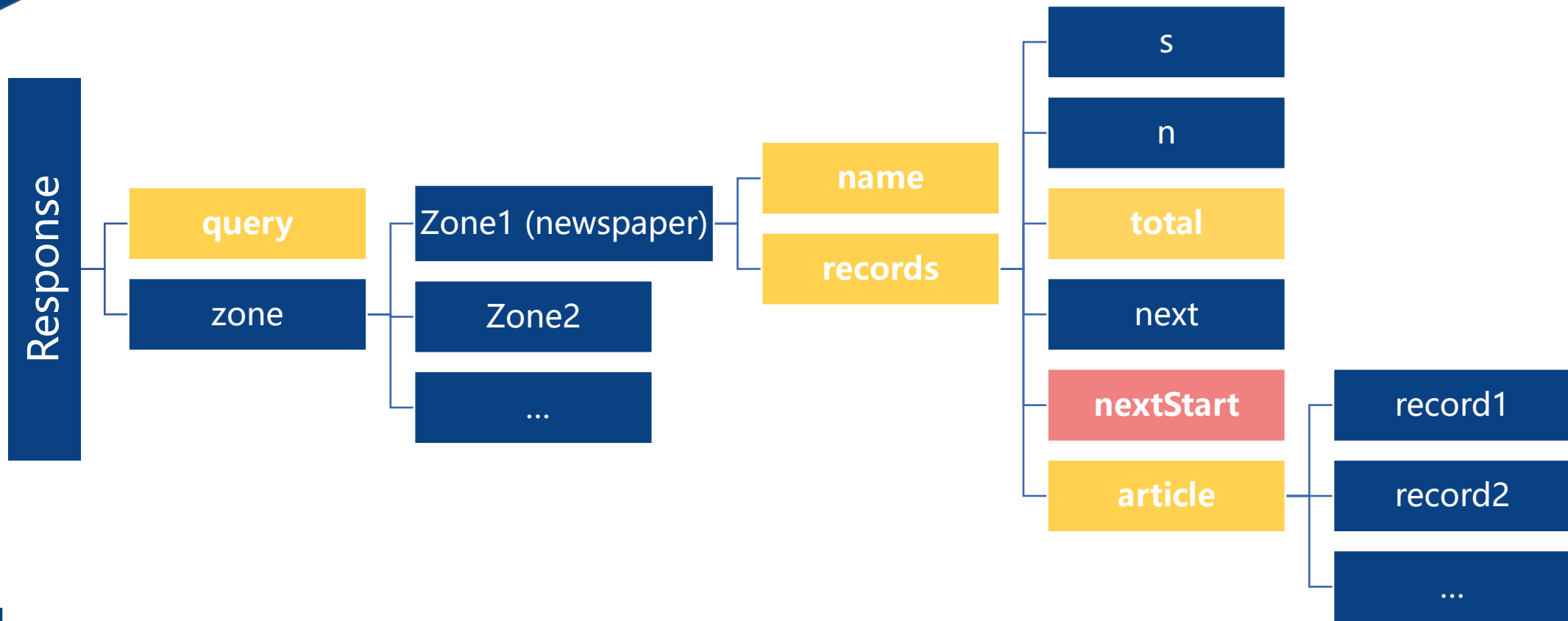
1. Select username

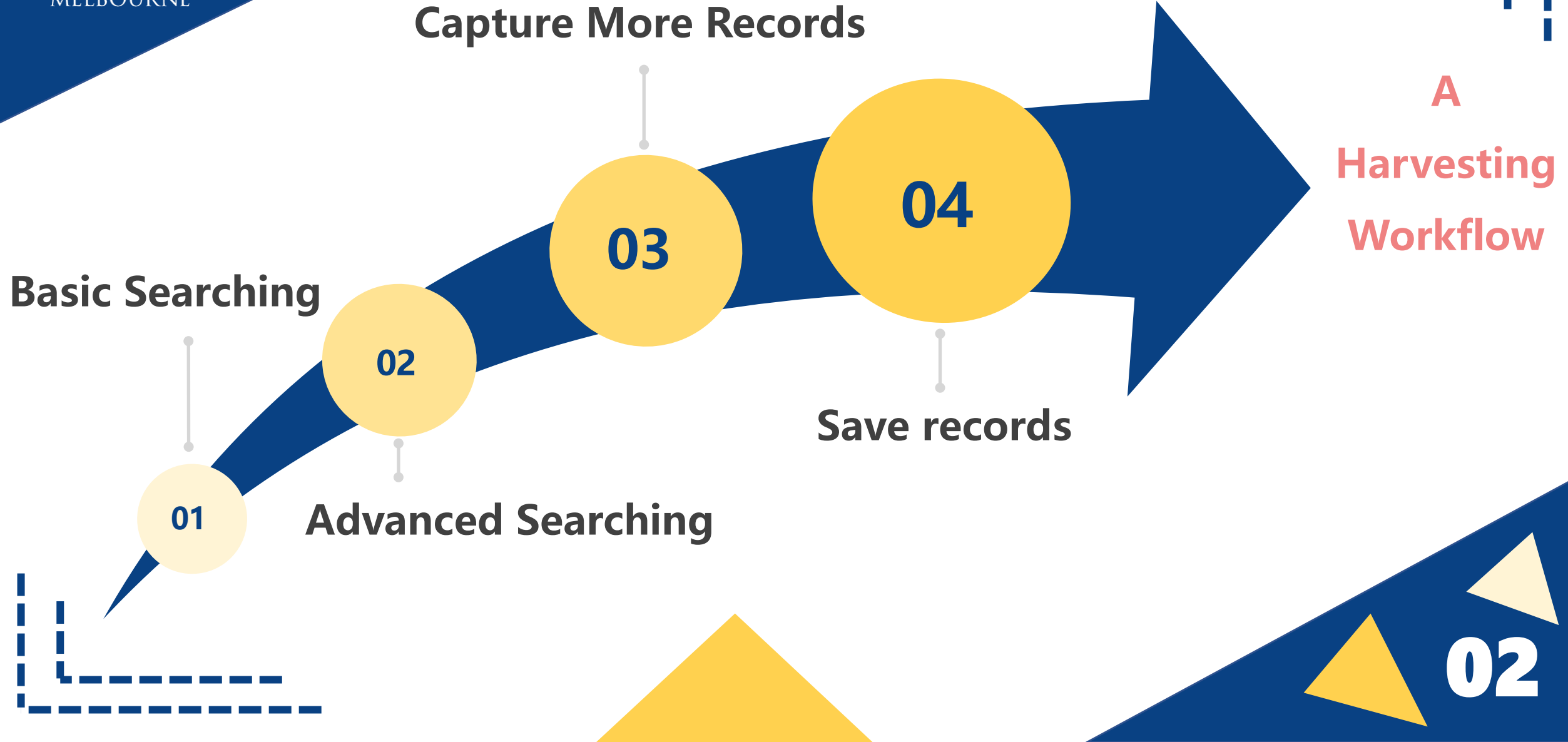
2. Select MY PROFILE

3. Select For Developers

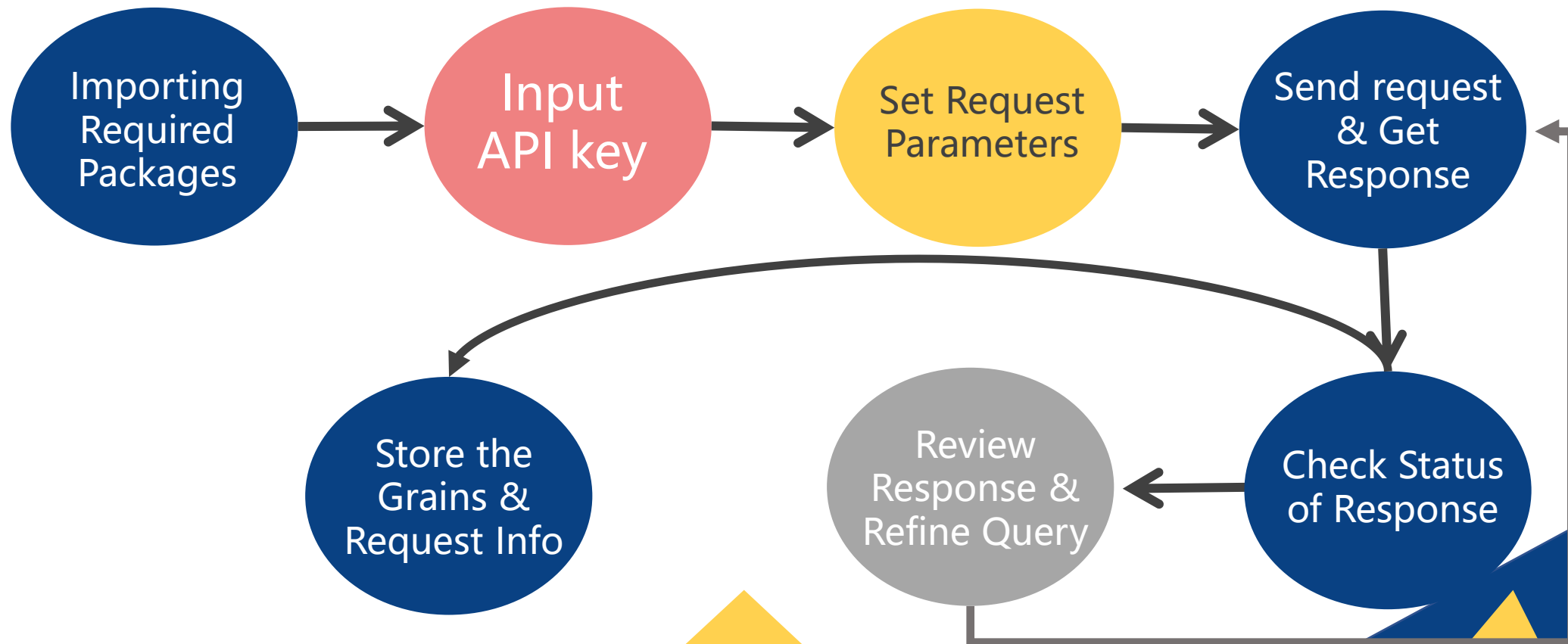


Structure of Trove Response

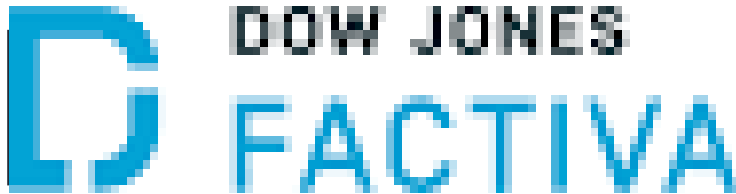




A Regular Workflow for Trove Harvesting




A Glance on




API

Factiva Database

 **DOW JONES FACTIVA**



HomeSearchAlertsNewslettersNews PagesCompanies/Markets



Free Text Search

Search Form

Examples

Query Genius  

0/2048

Date Duplicates

Search

Source

Author

Company

Factiva Expert Search

Subject

Industry

Region

Look up

Language

More Options

► All Sources

► All Authors

► All Companies


►

► All Subjects

► All Industries

► All Regions

►

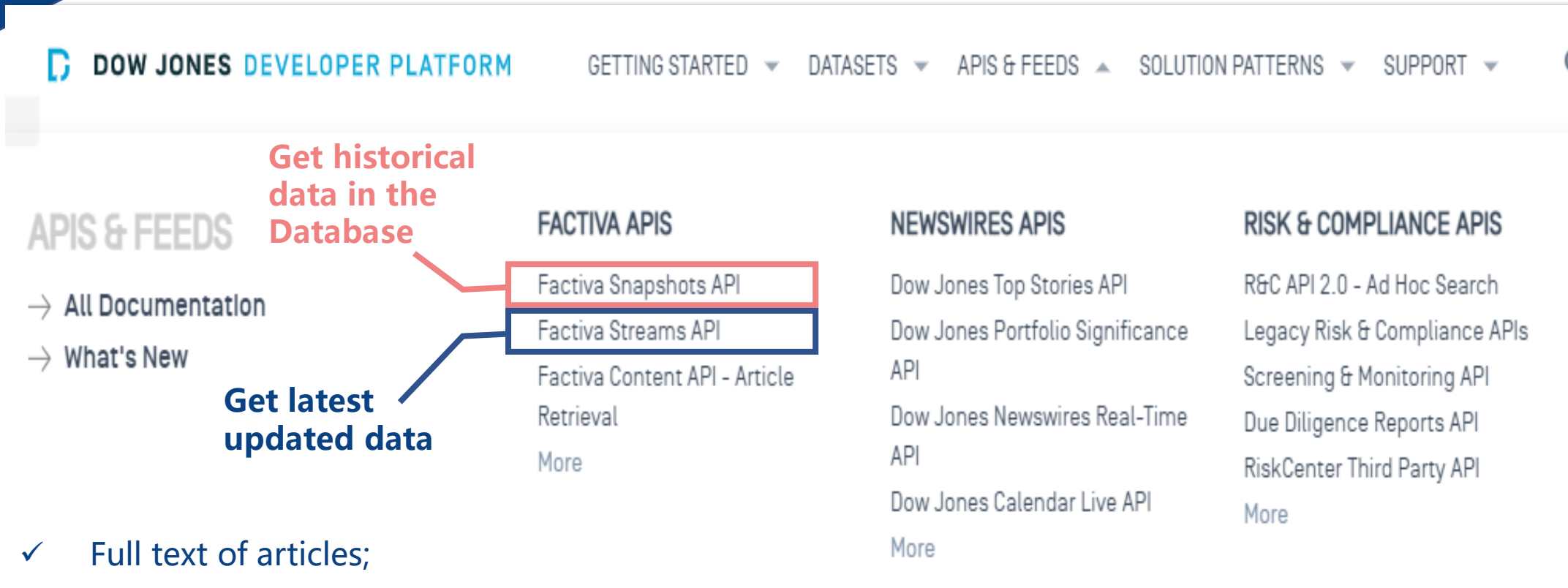
► English 

►

Clear Search

Search

Factiva APIs in Dow Jones Developer Platform



The screenshot shows the Dow Jones Developer Platform website. The navigation bar includes links for GETTING STARTED, DATASETS, APIS & FEEDS, SOLUTION PATTERNS, and SUPPORT. The 'APIS & FEEDS' section is highlighted, and a red arrow points to the 'Factiva APIs' section. The 'Factiva APIs' section lists the Factiva Snapshots API and Factiva Streams API, which are highlighted with a red and blue box respectively. A red arrow points to the 'Factiva Snapshots API' with the text 'Get historical data in the Database'. A blue arrow points to the 'Factiva Streams API' with the text 'Get latest updated data'. The 'NEWSWIRES APIS' and 'RISK & COMPLIANCE APIS' sections are also visible.

APIS & FEEDS

- All Documentation
- What's New

Get historical data in the Database

Get latest updated data

FACTIVA APIS

- Factiva Snapshots API
- Factiva Streams API
- Factiva Content API - Article Retrieval
- More

NEWSWIRES APIS

- Dow Jones Top Stories API
- Dow Jones Portfolio Significance API
- Dow Jones Newswires Real-Time API
- Dow Jones Calendar Live API
- More

RISK & COMPLIANCE APIS

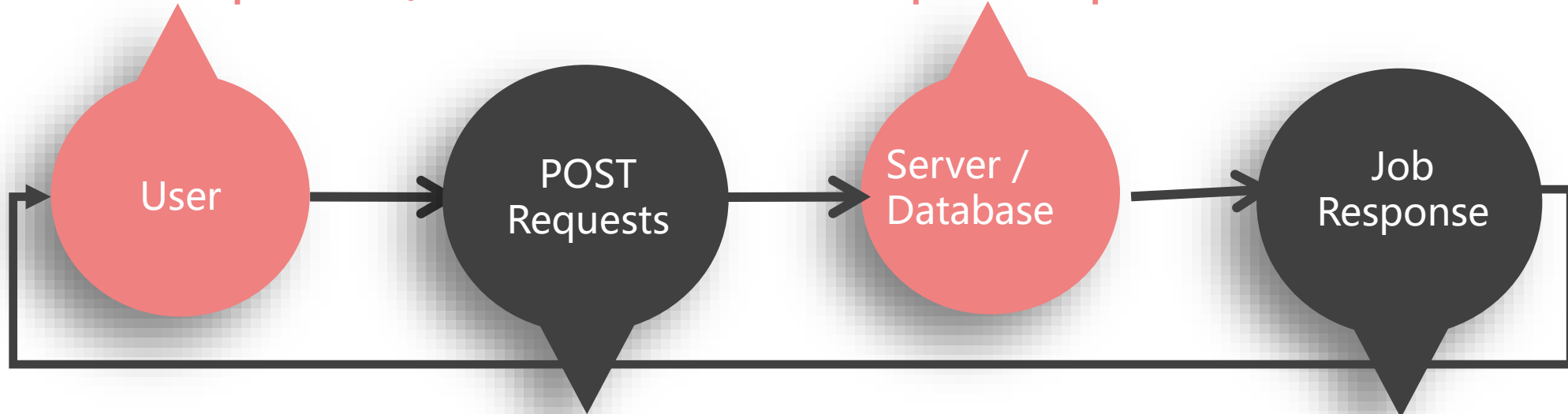
- R&C API 2.0 - Ad Hoc Search
- Legacy Risk & Compliance APIs
- Screening & Monitoring API
- Due Diligence Reports API
- RiskCenter Third Party API
- More

- ✓ Full text of articles;
- ✓ More Delivery formats (csv, json, xml, avro);
- ✓ Faster, safer, more advanced

How does Factiva Snapshot API Work?

"Hey, I want some English articles about covid-19 since September 1, 2020!"

Job State:
"Snapshot is queuing ...creating...running"
"Snapshot complete!"



Root URL + Request Body +
API key

JOB_STATE_DONE state
+ URIs to files that can
be downloaded

Types of Factiva Snapshots

Basic Snapshot

Provide URIs that can access to the files you can download

Snapshot Explains

Provide the number of documents that match your query

factiva-core & factiva-news
can help you do these
operations in a few lines.

Snapshot Updates

Provide updated information for a previously created Snapshot that has since been added to the Dow Jones corpus

Snapshot Analytics

Provide insight on the articles matching your query by breaking them down over periods and source publication

How to Construct a Query for Factiva Snapshot

Select Clause: select the returned fields you want

There are 40 fields provided in Factiva Snapshot API.

```
request_body = {  
  "query": {  
    "select": ["art", "action", "body", "byline", "credit",  
    "where": "language_code='fr' AND (publication_date >= '2002-04-02 00:00:00' AND publication_date < '2002-04-03 00:00:00')",  
    "includes": {"company_codes": ["prgml"]},  
    "excludes": {  
      "currency_codes": ["usd"],  
      "market_index_codes": ["xdjti", "xf500", "xnyci"],  
    },  
    "format": "json",  
    "limit": 10,  
  }  
}
```

Where Clause: filtered by value of fields

Include & Exclude Clause: filtered by value of "code fields"

Format Clause: control the format of returned files.

Limit Clause: control the number of returned records

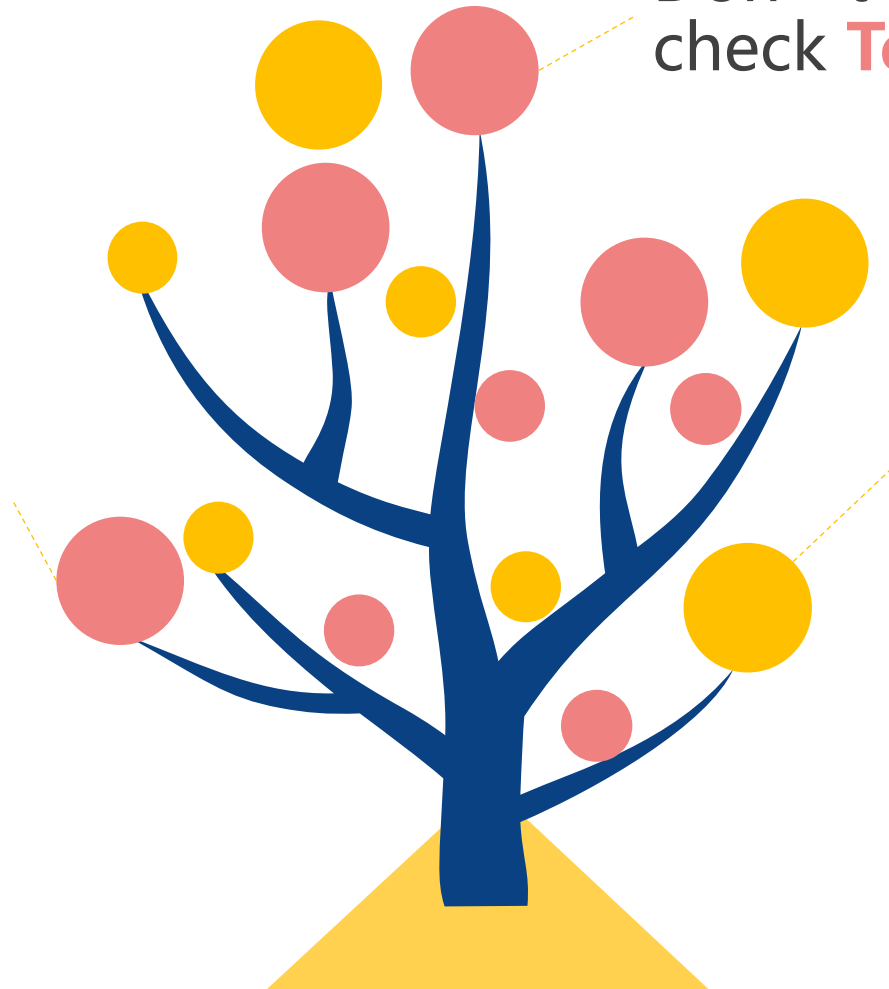
Tips in Using API

● ● ● ● ● Some Tips ● ● ● ● ●

Query construction
is more important
than **coding efficiency**.

Don' t forget to
check **Terms of Use!**

Make good use of
existing tools!



● ● ● ● ● Some References ● ● ● ● ●

- **Trove**

- Searching Skills in Trove ([basic](#), [advanced](#))
- [API technical guide](#)
- [Using the Trove API with Excel Spreadsheets](#)
- [Trove API Console](#) (no API key required!)
- [Trove newspaper & gazette harvester](#)

- **Factiva**

- [Factiva Snapshots API Official Documentation](#)
- [Factiva Sample Notebooks](#) (factiva-core + factiva-news)

Thanks for Listening

I acknowledge the Traditional Owners of the land on which I work, and
pay my respects to the Elders, past and present.

Zhixin (Ivy) Zheng
(ivy.zheng@unimelb.edu.au)