Standard Data Format

Semi-Structured Data

For semi-structured data consisting of text, there are a number of formats that predominate: Extensible Markup Language (XML) and JavaScript Object Notation (JSON) and others
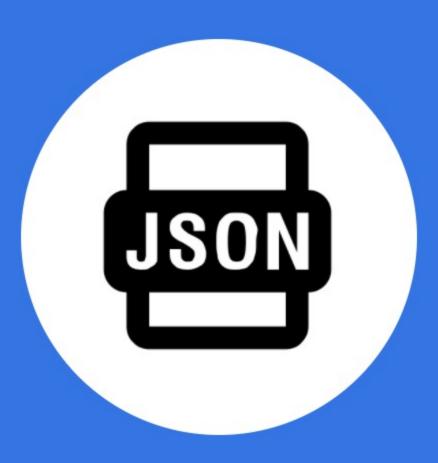
XML

1. Extensible Markup Language (XML)

In computing, Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable through use of tags that can be created and defined by users.

Source: https://en.wikipedia.org/wiki/XML

```
<?xml version="1.0">
<contact-info>
        <name>Teo Bee Guan</name>
        <company>Monash University Malaysia</company>
        <extension>46072</extension>
</contact-info>
```
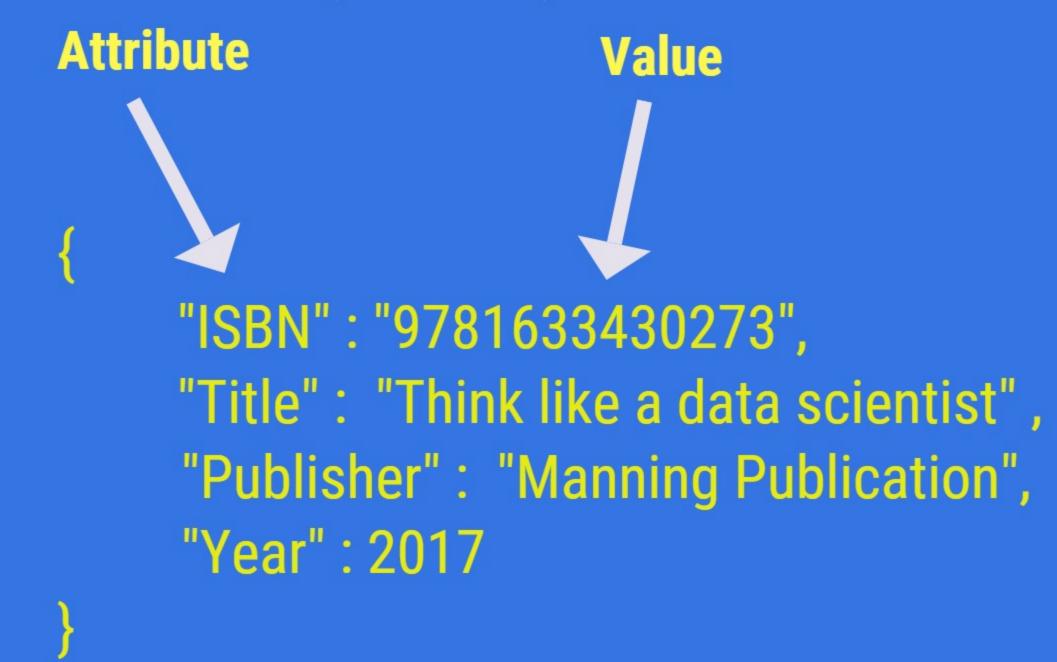
JSON

2. Javascript Object Notation (JSON)

In computing, JavaScript Object Notation or JSON is an open-standard file format that uses human-readable text to transmit data objects consisting of attribute−value pairs and array data types (or any other serializable value).

Source: https://en.wikipedia.org/wiki/JSON

Attribute   **Value**

```
{
    "ISBN" : "9781633430273",
    "Title" :  "Think like a data scientist" ,
    "Publisher" :  "Manning Publication",
    "Year" : 2017

}
```

YAML

YAML is a human-readable data serialization language. It is commonly used for configuration files, but could be used in many applications where data is being stored (e.g. debugging output) or transmitted (e.g. document headers).

https://en.wikipedia.org/wiki/YAML

```
application: guestbook
version: 1
runtime: python27
api_version: 1
threadsafe: true

handlers:
- url: /css
    static_dir: css

- url: /.*
     script: main.application

libraries:
- name: webapp2

version: latest
```