PMSI Scraping Estimate – Template

1. Requirements

1.1 Basic details

**Site Name**

|  |
| --- |
| E-leclerc. |

**Site URL (or example if multiple URLs)**

|  |
| --- |
| <http://www.e-leclerc.com/> |

**Country / Regional Requirements**

|  |
| --- |
| France |

**Objective**

|  |
| --- |
| Aim is to scrape all available store information once only. |

1.2 Scrape Timings

**Frequency**

|  |
| --- |
| Once |

**Timing**

|  |
| --- |
| N/A |

1.3 Output

|  |
| --- |
| Need data in SQL or exported into one excel file with consistent column ordering. |

1.4 Pilot (optional)

|  |
| --- |
|  |

1.5 Other

|  |
| --- |
|  |

1. Deadlines

**Estimate**

|  |
| --- |
|  |

**Pilot / Test data**

|  |
| --- |
|  |

**Production data**

|  |
| --- |
|  |

1. Fields Summary

**Fields that will be scraped:**

|  |
| --- |
| -- StoreID might available can be confirmed after first scrape (codePanonceau).  -- Postcode  -- Fax, Telephone  -- Latitude, Longitude  -- City name,  -- Street Name  Lots of others field available in Json Documents  \*\*\* One example of Json Document added at the end of the template |

**ID Field:**

|  |
| --- |
| * codePanonceau (can be confirmed after first scrape) |

**Comments:**

|  |
| --- |
| Direct postcode search doesn’t return results, need to make call by passing Latitude and Longitude.  Latitude has to be in 8 digits and need to remove “.” From the latitude and longitude to get store information.  Need core changes to clean up “.” From Latitude and Longitude and need to append 0 at the end of Latitude and Longitude or multiply Latitude and Longitude by 1000000 to remove “.” And add extra zero at the end. |

1. Estimates

**Development Time (Total):**

|  |  |  |  |
| --- | --- | --- | --- |
| Low: | 2 Days | High: | 2.5 Days |

**Development Time (Pilot):**

|  |  |  |  |
| --- | --- | --- | --- |
| Low: |  | High: |  |

**Scrape Time:**

|  |  |  |  |
| --- | --- | --- | --- |
| Low: | 1 | High: | 1 |

**Scrape Cost:**

|  |  |  |  |
| --- | --- | --- | --- |
| Low: |  | High: |  |

|  |  |  |
| --- | --- | --- |
| Site complexity: | 5 out of 10 |  |

**Comments:**

|  |
| --- |
| Json Document Example: The Josn document returns for each Latitude and Longitude.  {  "centrale": null,  "codePanonceau": "0706",  "coordonnee": {  "boitePostale": null,  "capitalSocial": null,  "codeInsee": null,  "codeNaf": null,  "codePostal": "06140",  "fax": "04.93.24.52.40",  "latitudeGPS": 43.72108,  "longitudeGPS": 7.109951,  "numTva": null,  "numeroVoie": "288",  "raisonSociale": null,  "siren": null,  "techid": 3073,  "telephone": "04.93.24.52.52",  "typeSociete": null,  "ville": "VENCE",  "villeRcs": null,  "voie": "AVENUE EMILE HUGUES"  },  "dateDelegation": null,  "delegationCentrale": null,  "estActif": null,  "fermetures": [{  "fermeture": {  "centrale": null,  "date": "2014-11-11T01:00:00",  "dateDebutAffichage": "2014-11-02T01:00:00",  "techid": 12010  },  "magasin": null,  "techid": 79790  }, {  "fermeture": {  "centrale": null,  "date": "2014-12-25T01:00:00",  "dateDebutAffichage": "2014-12-12T01:00:00",  "techid": 12011  },  "magasin": null,  "techid": 80384  }],  "horairesExceptionnels": null,  "horairesMagasins": [{  "codeJour": "2",  "fermetureApresMidi": 72000,  "fermetureMatin": null,  "heureFermetureApresMidi": 20,  "heureFermetureMatin": null,  "heureOuvertureApresMidi": null,  "heureOuvertureMatin": 8,  "magasin": null,  "minuteFermetureApresMidi": 0,  "minuteFermetureMatin": null,  "minuteOuvertureApresMidi": null,  "minuteOuvertureMatin": 30,  "offre": null,  "ouvertureApresMidi": null,  "ouvertureMatin": 30600,  "techid": 21506  }],  "nom": "VENCE",  "nomNormalise": "vence",  "offres": null,  "rattachementClubs": null,  "techid": 257  }, |