**Uitleg van Update\_observations.R:**

Dit script verwerkt de data van waarnemingen van rivierkreeften, handmatig gedownload vanuit waarnemingen.be. Het script volgt een aantal stappen om nieuwe data te identificeren, onjuiste locaties te corrigeren en data te filteren op basis van een aantal regels. Uiteindelijk wordt de bewerkte data opgeslagen in verschillende CSV-bestanden voor verdere analyse.

1. **Data-initialisatie**:
   * Het script controleert of er een eerdere dataset (last\_import.csv) beschikbaar is. Zo ja, dan wordt deze ingelezen; zo niet, wordt een lege dataset aangemaakt.
2. **Nieuwste data inladen**:
   * Het script zoekt naar CSV-bestanden in de map ./data/observations/ die overeenkomen met een specifiek patroon (event-craywatch-.\*.csv). Dit zijn de nieuwe exportbestanden van Natuurpunt.
3. **Data verwerken**:
   * Elk bestand wordt ingelezen, en de soortnaam wordt afgeleid uit de bestandsnaam. De soortnaam wordt als een nieuwe kolom (soort) aan het bestand toegevoegd.
   * De bestanden worden samengevoegd tot één grote dataset met alle nieuwe waarnemingen.
4. **Sorteren op locatie-ID**:
   * De data wordt gesorteerd op basis van de kolom Code.sample.location om consistentie te behouden.
5. **Nieuwe waarnemingen filteren**:
   * Waarnemingen die al eerder geïmporteerd zijn (in last\_import.csv), worden gefilterd uit de nieuwe dataset om alleen de unieke nieuwe waarnemingen te behouden.
6. **Correctie van foutieve locatie-ID's**:
   * Er wordt gecontroleerd op onjuiste formattering van de locatie-ID's (locID), zoals verkeerde hoofdletters of symbolen. Dit wordt gecorrigeerd zodat alle locatie-ID's voldoen aan een vast patroon (bijv. I\_0000\_0). Soms kan dit niet automatisch gebeuren, dan blijven deze incorrect staan.
7. **Filteren op opeenvolgende vangstdagen**:
   * Het script combineert de nieuwe data met eerdere datacheck-bestanden (datacheck.csv) om vangstlocaties te identificeren met vier opeenvolgende dagen van waarnemingen.
   * Vervolgens worden locaties gefilterd die voldoen aan het criterium van **exact** vier opeenvolgende vangstdagen.
8. **Opslaan van geverifieerde data**:
   * Data die voldoet aan het criterium van vier opeenvolgende dagen wordt toegevoegd aan een bestand met gevalideerde data (cleandata.csv).
   * De niet-gevalideerde data wordt opgeslagen voor toekomstige bewerkingen in datacheck.csv.
9. **Overschrijven van de importbestanden**:
   * De nieuwe gegevens van deze verwerking worden opgeslagen in last\_import.csv zodat de volgende keer alleen nieuwe waarnemingen worden toegevoegd.

**Schematische voorstelling:**

+---------------------------------+

| Natuurpunt CSV-bestanden |

| (event-craywatch-\*.csv) |

+---------------------------------+

|

v

+---------------------------------+

| Inladen en soortnaam |

| toevoegen aan dataset |

+---------------------------------+

|

v

+---------------------------------+

| Sorteer data op locatie-ID |

+---------------------------------+

|

v

+----------------------------------------------+

| Vergelijk met vorige import (last\_import) |

| en filter nieuwe waarnemingen |

+----------------------------------------------+

|

v

+------------------------------------------------+

| Correctie van foutieve locatie-ID's |

+------------------------------------------------+

|

v

+-------------------------------------------------+

| Filter voor vier opeenvolgende vangstdagen |

+-------------------------------------------------+

|

v

+------------------------------------+ +-----------------------------------+

| Niet-gevalideerde data (datacheck)| | Gevalideerde data (cleandata.csv) |

+------------------------------------+ +-----------------------------------+

| |

v v

Opslaan als datacheck.csv Toevoegen aan cleandata.csv

en laatste import opslaan

als last\_import.csv

Het script zorgt er dus voor dat nieuwe waarnemingen op de juiste manier worden verwerkt en gevalideerd voordat ze worden opgeslagen voor verdere analyse.