



ISTITUTO NAZIONALE DI **O**ceanografia e di **G**eofisica **S**perimentale

EMODNet Chemistry Beach litter data validation process

Cociancich A., Brosich A., Altenburger A., Geletti E.,
Chaves Montero M.D.M., Vinci M.,
Molina Jack M.E. e Partescano E.

| | |
|---|-----------|
| Introduzione | 3 |
| 1. Beach litter data flow | 3 |
| 2. Validazione | 5 |
| 2.1 Processo di validazione e verifica dei vincoli | 5 |
| 3. Caricamento | 9 |
| 4. Produzione di file di uscita | 9 |
| 5. Conclusioni | 11 |
| APPENDICE 1: Lista dei controlli eseguiti dal validatore di Beach Litter | 12 |

Introduzione

Nell'ambito della terza fase di EMODnet Chemistry, l'attenzione è stata rivolta alla raccolta, standardizzazione, controllo di qualità e condivisione di dati e prodotti di “*marine litter*”. Lo sviluppo di un sistema Europeo per la gestione dei dati di marine litter è stato colto come un'opportunità ma allo stesso tempo come una vera e propria sfida dovuta alla diversità di protocolli, formati ed ontologie utilizzati per la raccolta dei dati (¹Molina Jack et al, 2019; ²Addamo et al, 2018).

Per avviare il sistema di gestione dei dati è stato necessario, non solo strutturare un database, ma creare anche una serie di strumenti che permettessero di automatizzare il processo: dalla formattazione alla validazione ed infine l'inserimento dei dati.

Nel dettaglio, questo documento fornirà le specifiche tecniche sul processo di validazione, inserimento ed estrazione dei dati di litter marino acquisiti lungo le spiagge.

1. Beach litter data flow

Il flusso di lavoro seguito per la gestione dei dati di beach litter è schematizzato nell'immagine seguente (Figura 1):

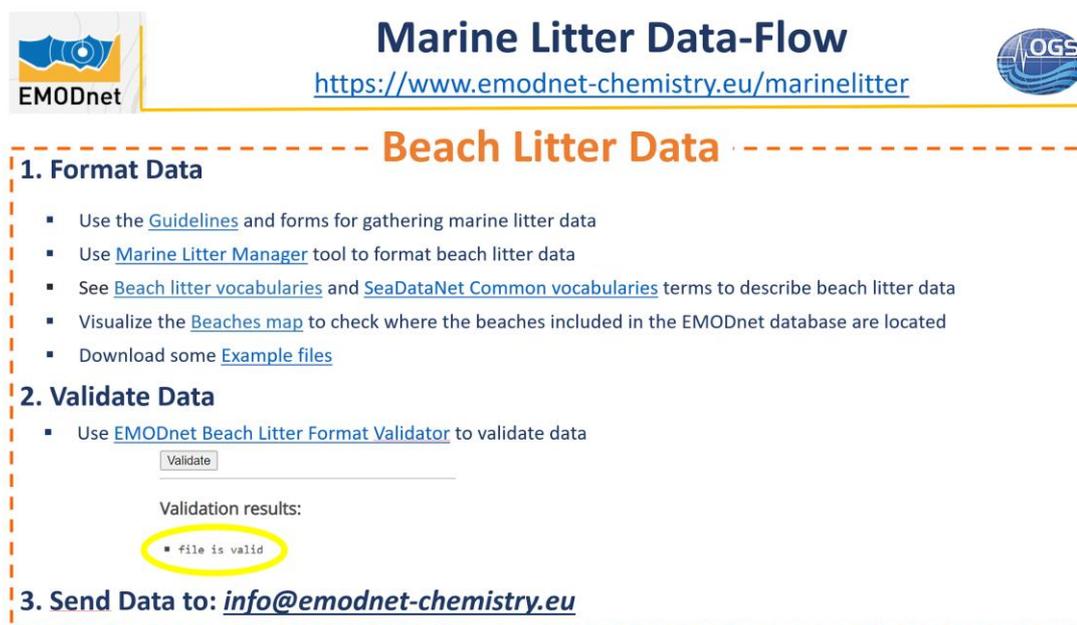


Figura 1. Schema del flusso di lavoro della gestione dei dati di beach litter

¹ Molina Jack, Maria Eugenia & Montero, Maria & Galgani, François & Giorgetti, Alessandra & Vinci, Matteo & Moigne, Morgan & Brosich, Alberto. (2019). EMODnet marine litter data management at pan-European scale. Ocean & Coastal Management. 181. 104930. 10.1016/j.ocecoaman.2019.104930.

² Addamo Anna Maria, Brosich Alberto, Chaves Montero Maria del Mar, Giorgetti Alessandra, Hanke Georg, Molina Jack Maria Eugenia, Vinci Matteo, Marine litter database: Lessons learned in compiling the first pan-European beach litter database, EUR 29469 EN, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-97866-1, doi:10.2760/68866, JRC112895.

1- Formattare dati seguendo le ³linee guida fornite dal progetto con particolare attenzione al:

- Formato dei metadati,
- Spiagge già catalogate
- Formato delle informazioni associate alle survey
- Formato standard per la descrizione degli animali

La base di dati EMODnet di Marine Litter prevede che i dati in ingresso abbiano un formato Excel formato da 4 fogli (beaches, surveys, animals, litter) secondo un formato predefinito.

Un esempio completo di dati in formato EMODnet lo si può trovare sul portale di EMODnet Chemistry al seguente indirizzo: (<https://doi.org/10.6092/a75ba101-ebb9-4bad-9b7f-423a1327c76f>).

2- I dati correttamente formattati vengono validati utilizzando il validatore *on-line* accessibile tramite interfaccia grafica (<https://www.emodnet-chemistry.eu/marinelitter/beachvalidator>), quest'ultimo si basa su un web service RESTful che permette di validare e caricare i dati all'interno del database e di estrarli in un formato Excel EMODnet compliant.

3- Se la validazione va a buon fine i dati vanno inoltrati via email a info@emodnet-chemistry.eu oppure a nodc@inogs.it

4- I dati validati vengono inseriti in banca dati tramite un comando in curl tramite una richiesta http ad un web service

5- Infine attraverso il web service si possono estrarre i dati nel formato di uscita seguendo il formato EMODnet descritto sopra.

Il formato dati EMODnet di Marine Litter è descritto nelle linee guida disponibili sul sito insieme al template e ad un esempio:

- [Guidelines and forms for gathering marine litter data](#)
- [Beach format template](#)

I dati provenienti dai vari partner sono idealmente ricevuti nel formato sopra definito. Altrimenti, è necessario produrre dei file nel formato EMODnet. Questa operazione di conversione può avvenire in modo manuale, in base al formato dei dati di partenza, oppure in modo semi-automatico utilizzando uno strumento in Python sviluppato per agevolare il processo di formattazione dei dati e convertirli nel formato EMODnet (<https://www.emodnet-chemistry.eu/marinelitter/manager>).

³ F. Galgani, A. Giorgetti, M. Le Moigne, A. Brosich, M. Vinci, M. Lipizer, M. E. Molina Jack, N. Holdsworth, R. Schlitzer, G. Hanke, G. Moncoiffe, D. Schaap, G. Giorgi, A. Addamo, M. D. M. Chaves Montero, 2020, Guidelines and forms for gathering marine litter data, 27/07/2020, 58 pp., DOI: 10.6092/15c0d34c-a01a-4091-91ac-7c4f561ab508

2. Validazione

Per accedere al validatore tramite la sua interfaccia grafica, si può utilizzare il seguente link:

<https://www.emodnet-chemistry.eu/marinelitter/beachvalidator>

Un'alternativa può essere accedere tramite REST Service, utilizzando una richiesta di tipo POST (realizzabile via browser solo tramite plugin appositi) ad esempio attraverso il comando *curl*.

Il nome del file può essere messo tra virgolette singole nel caso il nome contenga caratteri speciali.

Il risultato della validazione è un elenco in formato testuale di eventuali errori o avvisi: gli errori riguardano il formato del dato e bloccano l'inserimento in banca dati, mentre gli avvisi sono relativi al contenuto del dato e permettono di fare un controllo di qualità del dato prima dell'inserimento in banca dati. Nel caso non ci siano errori o avvisi l'output sarà il messaggio: *"File is valid"*.

Ad esempio:

Errori: Unable to convert field: Surveys.M54, value: WGS84 (4979), type: Short, exception: NumberFormatException - For input string: "WGS84 (4979)"

Avvisi: Surveys[25] - Survey's start coordinate "39.332203" - "23.026257" are pretty far from Beach one "39.333305" - "22.933918" (WARNING)

Dopo aver corretto gli errori evidenziati dalla validazione, i dati vanno copiati nella cartella "validated" sul server per il suo inserimento nel database.

2.1 Processo di validazione e verifica dei vincoli

Il processo di caricamento dei dati nel database passa attraverso diverse fasi a partire dai file in formato Excel.

- Controllo del numero e della posizione delle colonne e verifica del nome della colonna
- Lettura delle righe dei fogli e creazione di oggetti Java per la successiva fase di validazione
- Validazione secondo le specifiche "Bean Validation". Le classi degli oggetti Java che rappresentano le righe in ingresso sono opportunamente annotate.

Nel caso la validazione abbia avuto esito positivo il Web Service procede all'inserimento nella base dati.

L'elenco dei controlli eseguiti è il seguente (la lista completa è consultabile nell'Appendice 1):

★ Beach File

| |
|---------|
| ValidId |
|---------|

BeachOrSurveyCoordinate
ValidBeachPosition
ValidBeachSurveyDistance
ValidBeachSurveyLengthInterval
ValidCountryOriginator
ValidEdmerpCode
ValidEdmoCode
ValidItemCode
ValidSurveyContent
ValidSurveyLength
ValidVocabTerm
BeachAlreadyPresentException
DuplicatedCodeElementException
DuplicatedCoordsElementException
PossibleDuplicatedElementException
WrongColumnPositionException
ValidationException

★ **Beach**

BeachCode
BeachName
Country
BeachInfoAmmendment
FillingDate
FillingName
FillingPhone
FillingMail
FillingInstitute
UrbanizationDegree
ReferenceBeach
BeachWidthLow
BeachWidthHigh
BeachLength
BeachLatitude
BeachLongitude
CoordinateSystem
BeachBack
BeachBackOther
BeachBackDevelopment
DevelopmentDescription
PositionMeasurementDate

CurrentsDirection
WindsDirection
BeachOrientation
BeachMaterial
BeachTopography
Obstacles
Usage1
Usage1Seasonality
Usage2
Usage2Seasonality
Usage3
Usage3Seasonality
BeachAccess
BeachCleaningSeasonality
SeasonalityMonth
CleaningFrequency
OtherDescription
CleaningMethod
CleaningResponsible
Notes
Comments

★ **Survey**

BeachCode
SurveyCode
SurveyType
DataPolicy
SurveyDate
Originator
Collator
ProjectCode
SurveyStartLatitude
SurveyStartLongitude
SurveyEndLatitude
SurveyEndLongitude
CoordinateSystem
SurveyLength
SurveyWidth
Surveyor1Name
Surveyor1Phone
Surveyor1Mail
Surveyor2Name
Surveyor2Phone

Surveyor2Mail
TownName
TownDistance
TownPosition
TownPopolation
WinterTourists
SpringTourists
SummerTourists
AutumnTourists
FoodOutlets
FoodOutletsDistance
FoodOutletsSeasonality
SeasonalityMonths
FoodOutletsPosition
ShippingLaneDistance
ShippingLaneTraffic
ShippingLaneTypes
ShippingLanePosition
HarbourName
HarbourDistance
HarbourPosition
HarbourType
HarbourSize
RiverName
RiverDistance
RiverPosition
WasteWaterDischarges
WasteWaterDistance
WasteWaterPosition
LitterPresence
LastCleaning
WeatherConditions
WeatherConditionsOther
AnimailsFound
AnimalsNumber
SurveyCircumstances
SpecialEvents
Notes
Comments

★ **Animal**

SurveyCode

Borgo Grotta Gigante, 23/11/2020

| |
|--------------------|
| Animal |
| State |
| Sex |
| Age |
| Entanglement |
| EntanglementNature |
| Comments |

★ Litter

| |
|-----------------------|
| SurveyCode |
| LitterReferenceList |
| ItemCode |
| ItemName |
| ParameterOriginalName |
| NoItems |
| Notes |
| Comments |
| OtherItemCodes |

Le classi che compongono il modello di input sono nel package *inputmodel.beach*.

Tali classi ed i loro attributi sono opportunamente annotate con:

@SheetName, che contiene il nome letterale del foglio al quale sono associate

@DataCell, che identifica gli attributi che corrispondono ad una colonna del foglio Excel.

3. Caricamento

Il caricamento dei dati, che hanno previamente superato con successo la fase di validazione, viene fatto in modo analogo alla validazione da riga di comando, ma usando il comando curl.

Dopo il caricamento, i dati devono essere spostati nella cartella "inserted" sul server.

Attraverso il trigger `trigger_cdi_management_trg_f()` la tabella `litter_cdis` viene aggiornata automaticamente per tenere traccia delle operazioni eseguite specificando se si tratta di insert, update o delete.

4. Produzione di file di uscita

Il web service può essere usato per produrre file di uscita seguendo il formato EMODnet.

Il comando da eseguire è:

Borgo Grotta Gigante, 23/11/2020

| Richiesta | Path | Descrizione |
|----------------------------------|--|---|
| Per un determinato paese | <i>/surveys/<paese></i> | Dove hostname è la macchina dove gira il web service ed il paese viene indicato con il codice ISO (per esempio IT per Italia). |
| Per un determinato paese ed anni | <i>/surveys/<paese>/<anni></i> | Gli anni, se più di uno, vanno specificati separandoli con virgole |
| Intero database | <i>/surveys</i> | |
| Output per le baselines | <i>/baselinesdata/<LISTA></i> | È stato definito un output specifico che viene usato per il calcolo delle soglie di rifiuti marini a livello Europeo (⁴ Hanke et al, 2019). Per ottenere i dati delle baselines in formato trasposto ("wide"), per ogni singola lista |

Analogamente sul server emodnet-chemistry:

| Richiesta | Path | Descrizione |
|-------------------------|--|---|
| Singola survey | <i>/beach/survey/<surveyId></i> | |
| Aggregated collections | <i>/beach/collections/<dataPolicy>/<surveyTypes></i> | Dove dataPolicy è uno dei valori del vocabolario SeaDataNet L08 (http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=L08) e surveyTypes è un elenco di valori definiti specificamente per la gestione del beach litter (https://www.emodnet-chemistry.eu/marinelitter/vocab) separato da virgole (nel caso si voglia specificarne più di uno) |
| Output per le baselines | <i>/beach/baselinesdata/<LISTA></i> | Dove LISTA è la lista di riferimento che si desidera come output (i.e. OSPAR, TSG_ML, UNEP, UNEP_MARLIN, ITA) |

⁴ Hanke G., Walvoort D., van Loon W., Addamo A.M., Brosich A., del Mar Chaves Montero M., Molina Jack M.E., Vinci M., Giorgetti A., EU Marine Beach Litter Baselines, EUR 30022 EN, Publications Office of the European Union, Luxemburg, 2019, ISBN 978-92- 76-14243-0, doi:10.2760/16903, JRC114129.

5. Conclusioni

Il processo di validazione è fondamentale per assicurare l'uniformità nel formato dei dati e per garantire un minimo controllo di qualità che precede l'inserimento in banca dati.

Nel tempo, attraverso l'implementazione del validatore e l'incremento dei controlli effettuati da quest'ultimo, la qualità dei dati inseriti è decisamente migliorata. I controlli che in precedenza erano effettuati manualmente, con l'ausilio del validatore vengono ora eseguiti prima del processo di inserimento; in questo modo vengono inseriti in banca dati solo dati di buona qualità.

Inoltre, il miglioramento dei controlli eseguiti dal validatore comporta anche una riduzione delle operazioni di cancellamento ed aggiornamento della banca dati che venivano eseguite manualmente dagli operatori, riducendo così la possibilità di introdurre errori nel sistema.

APPENDICE 1: Lista dei controlli eseguiti dal validatore di Beach Litter

Beach File

| nome | tipo | entità | classe | campi | check |
|------------------------------------|-------------|------------------|--|--|--|
| ValidId | constraints | Beach, Survey, A | ValidId | AnimalCondition, AnimalSex, Country, | controlla che l'id (chiave esterna) esista |
| BeachOrSurveyCoordinate | constraints | Beach, Survey | BeachOrSurveyCoordinate | BeachLat, BeachLon, SurveyStartLat, SurveyStartLon, SurveyEndLat, SurveyEndLon, SurveyLength | esistenza una coppia di coordinate tra Beach/SurveyStart/SurveyEnd e SurveyLength oppure entrambe le coordinate della Survey |
| ValidBeachPosition | constraints | Beach | ValidBeachPosition | BeachLat, BeachLong | distanza delle coordinate inferiori a 0.005 dalla linea di costa |
| ValidBeachSurveyDistance | constraints | Beach, Survey | ValidBeachSurveyDistance | BeachLat, BeachLon, SurveyStartLat, SurveyStartLon | distanza delle coordinate survey da quelle beach entro un range |
| ValidBeachSurveyLengthInterval | constraints | Survey | ValidBeachSurveyLengthInterval | SurveyStartLat, SurveyStartLon, SurveyEndLat, SurveyEndLon | lunghezza survey entro un range |
| ValidCountryOriginator | constraints | Beach, Survey | ValidCountryOriginator | Country, Originator | controlla che il beach country sia lo stesso dell'originator (EDMO) |
| ValidEdmerpCode | constraints | Survey | ValidEdmerpCode | ProjectCode | controlla che il codice EDMERP esista |
| ValidEdmoCode | constraints | Survey | ValidEdmoCode | Originator, Collator | controlla che il codice EDMO esista |
| ValidItemCode | constraints | Litter | ValidItemCode | ItemCode | controlla che l'item code esista |
| ValidSurveyContent | constraints | Survey, Litter | ValidSurveyContent | LitterPresence, Noltems | la somma dei Noltems sia coerente con il valore di LitterPresence |
| ValidSurveyLength | constraints | Beach, Survey | ValidSurveyLength | BeachLength, SurveyLength | la lunghezza della Survey non sia più grande di quella della Beach |
| ValidVocabTerm | constraints | Beach, Survey | ValidVocabTerm | CoordinateSystem, DataPolicy | controlla l'appartenenza ad un vocabolario |
| BeachAlreadyPresentException | exception | Beach | BeachLitterLoader | BeachLat, BeachLon | verifica se c'è già una spiaggia con la stessa coppia di coordinate |
| DuplicatedCodeElementException | exception | Beach, Survey | BeachLitterImporter | SurveyCode, BeachCode | verifica se c'è già una spiaggia/survey con lo stesso codice |
| DuplicatedCoordsElementException | exception | Survey | Beach | SurveyStartLat, SurveyStartLon, SurveyEndLat, SurveyEndLon, SurveyDate | verifica se c'è già una survey con la stessa coppia di coordinate e data |
| PossibleDuplicatedElementException | exception | Litter | Survey | | segnala eventuali items duplicati |
| WrongColumnPositionException | exception | generico | BeachLitterController, BeachLitterImporter | generico | verifica la posizione delle colonne (in importazione) |
| ValidationException | exception | generico | BeachLitterController | generico | verifica se il file è valido |

Beach

| campo | tipo | vincoli |
|--------------------------|---------|----------------|
| BeachCode | String | NotBlank |
| BeachName | String | NotBlank |
| Country | String | NotBlank |
| BeachInfoAmmendment | Boolean | |
| FillingDate | Date | Past |
| FillingName | String | |
| FillingPhone | String | |
| FillingMail | String | Mail |
| FillingInstitute | String | |
| UrbanizationDegree | String | ValidId |
| ReferenceBeach | Boolean | |
| BeachWidthLow | Integer | Positive |
| BeachWidthHigh | Integer | Positive |
| BeachLength | Integer | Positive |
| BeachLatitude | Decimal | Min -90,0 |
| BeachLongitude | Decimal | Min -180,0 |
| CoordinateSystem | Integer | L10 |
| BeachBack | String | ValidId |
| BeachBackOther | String | |
| BeachBackDevelopment | Boolean | |
| DevelopmentDescription | String | |
| PositionMeasurementDate | Date | Past |
| CurrentsDirection | String | ValidId |
| WindsDirection | String | ValidId |
| BeachOrientation | String | ValidId |
| BeachMaterial | String | ValidId |
| BeachTopography | String | |
| Obstacles | String | UniqueElements |
| Usage1 | String | |
| Usage1Seasonality | Boolean | |
| Usage2 | String | |
| Usage2Seasonality | Boolean | |
| Usage3 | String | |
| Usage3Seasonality | Boolean | |
| BeachAccess | String | ValidId |
| BeachCleaningSeasonality | Boolean | |
| SeasonalityMonth | Integer | Min 1 |
| CleaningFrequency | String | ValidId |
| OtherDescription | String | |
| CleaningMethod | String | ValidId |
| CleaningResponsible | String | |
| Notes | String | |
| Comments | String | |

Survey

| campo | tipo | vincoli |
|------------------------|---------|-----------------|
| BeachCode | String | NotBlank |
| SurveyCode | String | NotBlank |
| SurveyType | String | NotBlank |
| DataPolicy | String | NotBlank |
| SurveyDate | Date | NotNull |
| Originator | Integer | ValidEdmoCode |
| Collator | Integer | ValidEdmoCode |
| ProjectCode | Integer | ValidEdmerpCode |
| SurveyStartLatitude | Decimal | Min -90 |
| SurveyStartLongitude | Decimal | Min -180 |
| SurveyEndLatitude | Decimal | Min -90 |
| SurveyEndLongitude | Decimal | Min -180 |
| CoordinateSystem | Integer | Positive |
| SurveyLength | Integer | Positive |
| SurveyWidth | Integer | Positive |
| Surveyor1Name | String | |
| Surveyor1Phone | String | |
| Surveyor1Mail | String | |
| Surveyor2Name | String | |
| Surveyor2Phone | String | |
| Surveyor2Mail | String | |
| TownName | String | |
| TownDistance | Decimal | PositiveOrZero |
| TownPosition | String | ValidId |
| TownPopolation | Integer | PositiveOrZero |
| WinterTourists | Integer | PositiveOrZero |
| SpringTourists | Integer | PositiveOrZero |
| SummerTourists | Integer | PositiveOrZero |
| AutumnTourists | Integer | PositiveOrZero |
| FoodOutlets | Boolean | |
| FoodOutletsDistance | Decimal | PositiveOrZero |
| FoodOutletsSeasonality | Boolean | |
| SeasonalityMonths | Integer | UniqueElements |
| FoodOutletsPosition | String | ValidId |
| ShippingLaneDistance | Decimal | PositiveOrZero |
| ShippingLaneTraffic | Integer | PositiveOrZero |
| ShippingLaneTypes | String | ValidId |
| ShippingLanePosition | String | ValidId |
| HarbourName | String | |
| HarbourDistance | Decimal | PositiveOrZero |
| HarbourPosition | String | ValidId |
| HarbourType | String | ValidId |
| HarbourSize | Integer | PositiveOrZero |
| RiverName | String | |
| RiverDistance | Decimal | PositiveOrZero |
| RiverPosition | String | ValidId |
| WasteWaterDischarges | Boolean | |
| WasteWaterDistance | Decimal | PositiveOrZero |
| WasteWaterPosition | String | ValidId |
| LitterPresence | Boolean | default true |
| LastCleaning | Date | Past |
| WeatherConditions | String | ValidId |
| WeatherConditionsOther | String | |
| AnimailsFound | Boolean | |
| AnimalsNumber | Integer | PositiveOrZero |
| SurveyCircumstances | String | |
| SpecialEvents | String | |
| Notes | String | |
| Comments | String | |

Animal

| campo | tipo | vincoli |
|--------------------|---------|----------|
| SurveyCode | String | NotBlank |
| Animal | String | NotBlank |
| State | String | ValidId |
| Sex | String | ValidId |
| Age | Integer | Positive |
| Entanglement | Boolean | |
| EntanglementNature | String | |
| Comments | String | |

Litter

| campo | tipo | vincoli |
|-----------------------|---------|----------|
| SurveyCode | String | NotBlank |
| LitterReferenceList | String | NotBlank |
| ItemCode | String | NotBlank |
| ItemName | String | |
| ParameterOriginalName | String | |
| NoItems | Integer | NotNull |
| Notes | String | |
| Comments | String | |
| OtherItemCodes | | |