D2.1

Data Processing Strategy

WP n° and title	**WP2 – Ecosystem services from the BCS: Demand analysis and the legal-political context**
Responsible Author(s)	**VLIZ**
Contributor(s)	**UGent-STEN, UGent-GhenToxLab, UAnt**
Dissemination level PU = Public; PP = Restricted to other program participants; RE = Restricted to a group specified by the consortium; CO = Confidential, only for members of the consortium	

DOCUMENT INFORMATION

Project Title	**SUMES: Sustainable Marine Ecosystem Services**
Status (F: final; D: draft; RD: revised draft):	**Final**
Planned delivery date	**31/01/2021 (M4)**
Actual delivery date	**18/02/2021 (M5)**

DOCUMENT HISTORY

Version	Date (MM/DD/YYYY)	Description of changes	Contributors
01	**09/02/2021**	**Set up Data Processing Strategy**	**VLIZ, UGent-STEN**
02	**16/02/2021**	**Overall feedback**	**UGent-STEN, UGent-GhenToxLab, UAnt**

TABLE OF CONTENTS

LIST OF FIGURES

Figures

Acronyms

BCS	**Belgian Continental Shelf**
ES	**Ecosystem Services**
MDA	**Marine Data Archive**
VLIZ	**Flanders Marine Institute**
WP	**Work Package**
DPS	**Data Processing Strategy**
DMP	**Data Management Plan**

1. Executive summary

Significant amounts of data and information are necessary to create an accurate model of ecosystem services (ES) that can reliably inform a sustainable management of the maritime space. To assess the conditions of ES from the Belgian Continental Shelf (BCS), data and information on parameters and indicators that have been identified as being decisive or indicative of changes in ES will be collected. For each of these parameters, availability will be checked within national and international databases, scientific and grey literature, project and technical reports, and other types of existing data sources. Once collected, data and information will be stored on the Marine Data Archive (MDA) of VLIZ, bringing together all inventory in a centralized data system. During storage, a data processing strategy will be developed and implemented. The processing strategy should ultimately assist in establishing a collection of good quality datasets that, even coming from a wide variety of data sources, are well usable and interoperable due to their uniform format.

2. Introduction

This document describes the data processing strategy for the SUMES project and describes the data flow, from the initial data needs, to the collection and archiving of the data. The Data Processing Strategy (DPS) makes clear how searches for data are organized and by whom, and how these collected datasets are handled before archiving of the data.

The DPS is an integral part of the SUMES Data Management Plan (DMP). The DMP describes everything necessary to make the research data discoverable, accessible, usable, and understandable in the long term: organising, documenting, archiving, cataloguing, and sharing (e.g. deliverables, geographical maps, conceptual schemes, etc.). The DMP is available at https://dmponline.be/. Access is restricted to SUMES partners.

3. Data needs

Data needs for SUMES are summarized in the project proposal as "parameters that have been identified as decisive or indicative for ES: physical, biological, chemical and geological properties as well as parameters related to human activities and demand for ecosystem services."
The Flanders Marine Data Centre (VMDC VLIZ) together with UGent-STEN, UGent-GhenToxLab and UAnt will actively search for these different types of data. To facilitate this process, an online editable excel sheet has been made available by VLIZ where every research institute can list their specific data needs.
The parameter list is available at:
https://docs.google.com/spreadsheets/d/1lBSc9cGjRndbbG_crBbsMb4Vdp4r_1o1mO1S9EWNNJc/edit#gid=0

The excel list has the following columns:

- Assessment: data context
- Group: institutional research group that requests/needs the data
- Datatype: describes the type of data looking for
- Parameter: name of the parameter and unit and if needed desired data format
- Area: specific area/region for data to cover
- Data owner/data source: possible data source or data holding
- Contact name data owner/data source: name of the person that should be contacted to provide the desired data
- Email data owner/ data source: email of contact to deliver the data
- Who: who will contact the data owner to submit their data?
- Status: current status of the request (Open= no action yet at this moment, in progress= data search started and ongoing and Done= data archived in the MDA in the correct SUMES folder)

4. Data Processing Strategy

4.1. Scheme

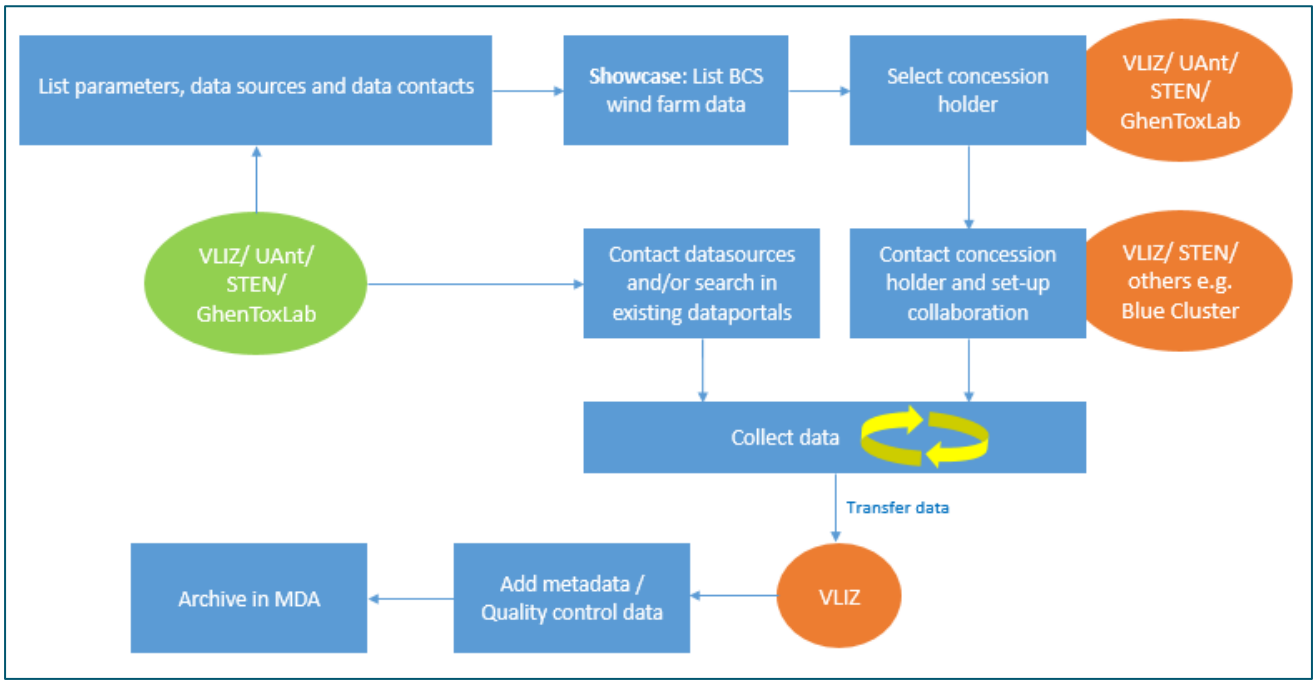

Figure 1: SUMES DPS

4.2. Data collection

The parameter list (to be found here) serves as a template to actively search for data. Data sources availability will be checked within national and international databases, scientific and grey literature, project and technical reports, and other types of existing data sources. For already available datasets, it is important to mention who will contact the data source for each specific parameter and therefore will be responsible for the data request. This ensures that double requests to data holders are avoided, which could slow down the data collection process. The task of data collection and storage is seen as an iterative process throughout the duration of the project. Progression in the project might identify different or new needs for data and information, as an addition to that already gathered. Moreover, it is also expected that some of the data needs will not be filled in and thus are identified as data gaps. Solutions will be provided in terms of metadata or close approximations to deal with eventual data gaps.

The following VLIZ data portals will be accessed:

- LifeWatch
- European Marine Observation and Data Network (EMODnet)
- Integrated Marine Environmental Readings & Samples (IMERS)
- World Register of Marine Species (WoRMS)
- Integrated Carbon Observation System (ICOS)

The following data providers will be contacted:

- Belgische statistiekbureau (Statbel)
- Statistical office of the European Union (Eurostat)
- Food and Agriculture Organization (FAO)
- European Market Observatory for Fisheries and Aquaculture Products Visveiling (EUMOFA)
- International Council for the Exploration of the Sea (ICES)
- Geofish
- Agentschap voor Natuur en Bos (ANB)
- Dredging, Environmental and Marine Engineering NV (DEME)
- Instituut voor Landbouw-, Visserij- en Voedingsonderzoek (ILVO)
- Wetenschappelijke Dienst Beheerseenheid van het Mathematisch Model van de Noordzee (BMM)

- Provincie West-Vlaanderen
- Royal Belgian Institute of Natural Sciences (RBINS)
- Federale Overheidsdienst Economie (FOD Economie)
- Belgian marine research groups (MOGs)
- Other…

For data specific to offshore wind farms, a general overview has been made (to be found here) on the availability and accessibility of data regarding existing wind farms in the BCS. The excel sheet is a living document that will be updated during the project and will be completed later on with specificities for complex cases. Based on this excel sheet, a concession holder will be selected and will be contacted in a separate process where, amongst others, STEN and Chantal Martens (Blue Innovation Officer at VLIZ) will have a discussion on data availabilities for the SUMES project and set up a data agreement on data access, restrictions and sharing of the data.

Eventually, VLIZ acts as data manager and will collect all datasets.

4.3. Metadata

Metadata will be added to every dataset and mapped to available standards:

- Data source (or data provider)
- Data contact
- Condition of use (license and restrictions, citations)
- Datatype
- Summary
- Info on parameters, units
- Info on sampling (protocol)
- Info on taxonomy
- Temporal coverage
- Spatial coverage
- Other…

4.4. Quality control of data

Firstly, this strategy will include quality control checks (use of standard vocabularies such as Marine Regions, WorMS, etc.) that assess and increase the dataset quality, a key step prior to usage in the other tasks and WPs. The quality of a dataset will be based on its completeness, uniqueness, timeliness, validity, accuracy and consistency. Secondly, the data processing strategy will contain an agreed upon data format in which the collected datasets will be stored. Where necessary, gathered and quality controlled datasets will be converted to this fixed format.

4.5. Archive in MDA

When the datasets are processed and considered as final, they will be stored in the SUMES project folder on the MDA. This is a secured folder: only the SUMES scientific partners (STEN, GhenToxLab, UAnt and VLIZ) have access to the files. Every partner has its own subfolder and, within these, extra folders were created based on the type of data (see figure 2).

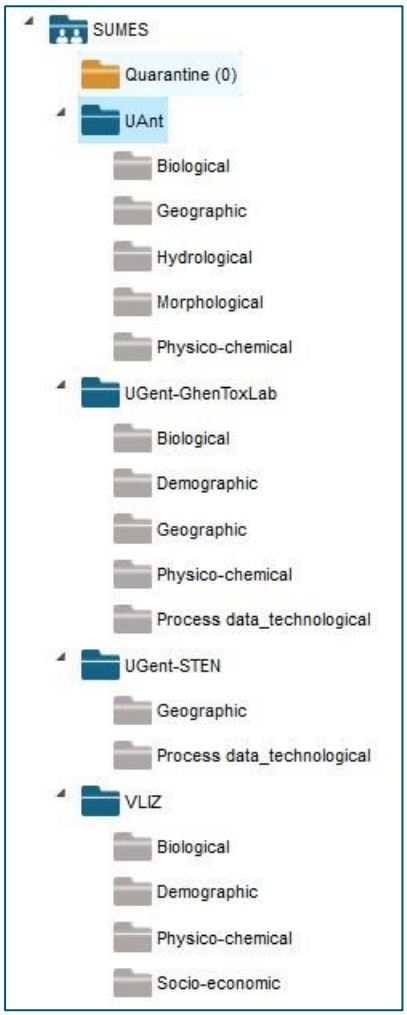

Figure 2: SUMES MDA folder structure

5. Conclusions

The SUMES DPS is a living document and serves as a roadmap on how data will be treated during the course of the project. It makes clear i) what data are needed, ii) where data can be found, iii) who will be the responsible for each data request, iv) how data will be formatted and v) where datasets eventually will be archived for future access and use.