BLUE MISSION BANOS

Supporting the Mission Ocean Lighthouse in the Baltic and North Sea Basins

Recommendations on data formats for KPIs and other data sources

Executive Summary

This brief provides a structured approach for **identifying and collecting data sources essential for monitoring Key Performance Indicators (KPIs)** in the context of a sustainable blue economy. It draws on the experience and lessons learned from the BlueMissionBANOS project, which aims to foster innovation and sustainability across sectors such as aquaculture, energy, waterborne transport, ports, and multi-use maritime activities. The guide emphasises the importance of **selecting data sources that are reliable, accessible, and suitable for automated collection**, thereby ensuring efficient integration and high-quality reporting. It outlines evaluation criteria, preferred data formats, retrieval methods, and representation techniques, offering practical examples to support

collaborators in their data collection efforts.

Preferred data acquisition methods highlighted in the guide include API protocols and structured file formats such as CSV and JSON, which allow for **automated, scalable, and accurate data collection**. The brief advises against the use of PDFs and other unstructured formats due to the challenges they pose for automation and reliable data extraction. By following these recommended practices, collaborators can contribute to a robust and efficient monitoring framework that enhances the accuracy and reliability of KPI measurement, supports evidence-based decision-making, and strengthens the capacity to track progress towards a sustainable blue economy.

Introduction

Selecting the right data formats is crucial for efficient and accurate data collection and integration. Here we outline the preferred methods for sharing data, highlighting their benefits and providing examples to help those with little or no knowledge of file formats understand and recognize them.

The preferred ways to share data are by providing API protocols, sharing CSV or JSON files. While API is not a data format, it is a highly effective way to share and access data.

Additionally, although CSV and JSON files are excellent choices, it is mandatory to maintain a pre-defined format to facilitate automation. Otherwise, manual intervention will be required to bring the data into a usable format, which can lead to issues such as distinguishing between first and last names.

It is also worth mentioning that PDFs, docx, and odt are not within the preferred data formats, as they can complicate the automation process.

Evaluation criteria for data sources

Criteria	Description
Initial Assessment	Identify potential data sources, verify credibility, and check accessibility.
Data Source Availability	Check if data is retrievable via API, CSV, JSON, or other formats.
Data Collectability	Assess if the data is collectable and ready for use.
Defining the Reference Point	Use the start of Mission Ocean as the initial reference point.
Defining the End Point	Establish specific, measurable goals.
Utilizing Existing Data and KPIs	Leverage existing datasets and KPIs.
Technological Readiness Level	Assess the maturity and feasibility of technologies.
Grouping Multiple Indicators	Incorporate various environmental, economic, and social indicators.

BLUE MISSION BANOS

① API

WHAT IS IT?

- An API (Application Programming Interface) is a **set of rules and protocols that allows different software applications to communicate with each other**. It's like a waiter in a restaurant who takes your order (request) and brings back your food (response) from the kitchen (server)!

BENEFITS

- **Automated Data Retrieval**: APIs allow for automated data retrieval, eliminating the need for manual data entry. This saves time and reduces the risk of errors.
- **Real-Time Updates**: APIs can provide real-time data updates, ensuring that the information is always current.
- **Scalability**: APIs can handle large volumes of data efficiently, making them suitable for scaling up data collection efforts.
- **Integration**: APIs can be easily integrated into existing systems and applications, facilitating seamless data flow.

EXAMPLE

- To retrieve weather data for a specific location, you can use a weather API instead of manually checking a website. You receive current weather information in a structured format that can be automatically processed and displayed in your application.

② CSV

WHAT IS IT?

- CSV (Comma-Separated Values) is a simple **file format used to store tabular data**, such as a spreadsheet or database. In a CSV file, each line is a data record, and each record consists of fields separated by commas.

BENEFITS

- **Simplicity**: CSV files are easy to create and read, making them accessible for users with minimal technical knowledge.
- **Compatibility**: CSV files can be opened and edited using common software like Microsoft Excel or Google Sheets.
- **Structured Data**: CSV files maintain a structured format, with each row representing a record and each column representing a field.

EXAMPLE

- A CSV file might look like this:

Name,Age,City\nAlice,30,New York\nBob,25,Los Angeles`
Name,Age,City

- In this example, each line represents a person, with fields for their name, age, and city.

BLUE
MISSION
BANOS

3 JSON

WHAT IS IT?
- JSON (JavaScript Object Notation) is a **lightweight data interchange format that is easy for humans to read and write and easy for machines to read and generate**. JSON represents data as key-value pairs and arrays.

BENEFITS
- **Human-Readable**: JSON is easy to read and understand, making it accessible for users with minimal technical knowledge.
- **Machine-Readable:** JSON is easily parsed by programming languages, making it ideal for automated data processing.
- **Flexibility**: JSON can represent complex data structures, including nested objects and arrays.

EXAMPLE
- Here is a what a JSON file might look like:

[{"Name": "Alice", "Age": 30, "City": "New York"}, {"Name": "Bob", "Age": 25, "City": "Los Angeles"}]

- In this example, each object represents a person, with keys for their name, age, and city.

4 PDF

WHAT IS IT?
- PDF (Portable Document Format) is a file format developed by Adobe that presents **documents in a consistent manner across diffe-rent devices and platforms**. PDFs are commonly used for **sharing documents that need to maintain their formatting**, such as reports, forms, and brochures.

CONSIDERATIONS
We do not recommend PDF as a data source for the following reasons:
- **Manual Extraction**: Data in PDFs often requires manual extraction, which is time-consuming and prone to errors. Automated extraction tools exist but can be unreliable and require additional processing.
- **Formatting Issues**: PDFs are designed to maintain formatting, not to facilitate data extraction. This can lead to difficulties in accurately extracting and structuring the data.
- **Lack of Standardization**: PDFs do not have a standardized structure for data, making it challenging to automate data collection and integration.

Those remarks also apply to docx and odt formats (Word and Open / Free Office).

EXAMPLE
A PDF document may contain a table with rows of numerical data. To use data from the PDF document, one would need to manually copy the table into a spreadsheet or use a PDF extraction tool, which may not accurately capture the data, especially column names and any merged columns. This manual process is more cumbersome, inefficient, and error-prone than obtaining structured data in formats like CSV or JSON.

Conclusion

Adhering to the guidance presented in this User Guide ensures a **standardized and efficient approach to data collection, enabling accurate and timely measurement of KPIs**. The principles outlined — from selecting automation-ready data sources to applying structured formats — provide a foundation for **robust monitoring frameworks and informed decision-making.**

Building on the experience of the BlueMissionBANOS project, these practices support the wider objective of advancing a **sustainable, innovative, and resilient** blue economy across maritime sectors.