

The Data Lifecycle - a Key Asset for the CLIMOS Project



Sergio Natal

Data Scientist | Predictia Intelligent Data Solutions



Due to the advancement of emerging technologies and the advent of Big Data as a fundamental part of modern life, there have been significant changes in how data is collected and analysed. As it happens to the life cycle of the sand fly species *Phlebotomus papatasi*, data also undergoes a detailed process of development and evolution to extract valuable information that can support decision-making.

A vast amount of data is managed in the CLIMOS project, spanning from sand fly occurrences to environmental information such as land use and climatic data including temperature, precipitation or soil moisture.

The wide variety of available data demands the implementation of

standard techniques and procedures to ensure the accuracy of the results obtained. In this context, CLIMOS focuses on aspects such as the use of official sources and the application of the FAIR principles. The use of open data facilitates CLIMOS in meeting these principles, ensuring that project results are available, accessible, interoperable and reusable.

For effective decision-making, it is not enough to have a large amount of data - it is essential to have quality data. Therefore, it is imperative to define the data life cycle properly, which involves addressing the following four main stages accurately:

- **Compilation:** The acquisition of necessary data for the project.

- **Storage:** At this maintenance and processing stage, data is organised, processed, and continuously cared for to keep it accessible and optimized for users. During this phase, processes such as integration, cleansing, and extract-transform-load (ETL) may be applied. This is one of the most crucial stages as homogenising and processing data into a common format ensures consistency and robustness, preparing them for subsequent analysis or modeling stages.
- **Usage:** Application of analysis and modeling using techniques such as niche models and machine learning, benefiting from a proper implementation of the storage stage.
- **Reuse or refuse:** It is essential to recognise the iterative nature of the data life cycle since the information provided by data can be reused for other projects. This implies that the last stage of the data life cycle can be the beginning of a new cycle, hence its continuous nature.

A proper implementation of each stage empowers CLIMOS to effectively manage, utilise, and reuse its data, maximizing its value and ensuring the quality of information to support informed decisions. Furthermore, this approach contributes to maintaining the integrity and security of data throughout its life cycle, acting as a solid protective mechanism and minimizing risks. CLIMOS exemplifies this approach by generating a homogeneous and robust dataset on the sand fly, combining various data sources to advance knowledge in this field and future projects.

Follow us:

