

Biodiversity 4.0: Standardizing biodiversity protocols for the private sector

Andre L. Acosta[‡], Valeria Tavares[‡], Guilherme Oliveira[‡], Leonardo C. Trevelin[‡], Ronnie Alves[‡], Nubia Marques[‡], Tereza Cristina Giannini^{‡,§}

[‡] Instituto Tecnológico Vale, Belém, Brazil

[§] Universidade Federal do Pará, Belém, Brazil

Corresponding author: Tereza Cristina Giannini (tereza.giannini@itv.org)

Abstract

Ensuring the preservation of biodiversity is essential for humankind, as the ecosystem services it provides are directly linked to human well-being and health. The private sector has increasingly recognized the need to achieve Environmental, Social, and Corporate Governance (ESG) through measurable indicators and effective data collection (Rashed 2021).

Extensive field research is often needed for private sector initiatives to generate socio-economic and environmental assessments, which usually requires hiring service providers. Regarding environmental and biodiversity information collections, the wide variety of data requires service providers to be specialized in many types of information, and therefore able to collect data on fauna and flora, soil and its microorganisms, genetic and evolutionary data, monitoring of the climate, conservation, and restoration areas, among many others.

Long-term monitoring, a generally common demand for the private sector (e.g., Shackelford (2018)), also relies on collecting various types of data often surveyed, gathered, and stored in a non-standardized fashion.

The lack of data standardization makes it difficult to integrate information into central databases (Henle 2013), creating a new demand to extract and convert data from different reports, which is often time and energy-consuming, and cost-ineffective. This task is generally conducted by non-specialists and may result in misinterpretation and digitization failures, compromising information quality.

The digital standardization of data is a key solution for solving these problems (Kuhl 2020), increasing efficiency in the collection, curation, and sharing of data, improving the quality and accuracy of the information, and reducing the risk of misinterpretation. The primary advantage is that the same professional who collects the data will digitize it into a common

database. The direct population of raw information into the database eliminates intermediate data conversion steps optimizing quality.

Here, we propose to generate a protocol for data collection in our institution (from the field, labs, museums, herbaria). This protocol is based on consolidated data standards, namely the [Darwin Core](#) (DwC). DwC is a glossary of terms that aims to standardize biodiversity information, which enables sharing data publicly. However, we are also creating new customized terms, classes, and respective metadata, such as species interaction, primarily to meet our need for long-term monitoring and assessments that are not covered by standard repositories.

To assess the types of surveyed and stored data required, we are interviewing biodiversity researchers from diverse scientific backgrounds about their specific data needs and the definitions of their recommended terms (metadata). Using this method, we aim to involve people in the development process, creating a more inclusive data protocol, ensuring that all possible data demands are covered, making the protocol more likely to be generally accepted.

Based on our interviews, one of the main difficulties in using a standardized glossary of terms is many unnecessary or unfillable data. This results from the search for comprehensiveness that also generates excessiveness. Taking this into account, we created a modular logic, selecting the best set of data (from a complete standardized database) for the specific demand or use.

For example, if this standard database is used to guide a floral survey, it will most likely not require variables on fauna, caves, hydrology, etc. In this way, the system exports a perfectly customized digital spreadsheet containing the variables that the research team wants to collect, but also recommending other variables of interest that can be obtained during fieldwork, increasing the efficiency and scope of the activity (which may be financially onerous).

We intend to make the system compatible with mobile technologies to be used indoors and outdoors, transferring the information directly to a virtual and integrative database. These open data collection protocols could be freely applied in other communities e.g., public research institutions, researchers' fieldwork, and citizen science projects.

We want our framework to be [FAIR](#), making our data more Findable, Accessible, Interoperable, and Re-usable, and will integrate the Internet of Things (IoT), Artificial Intelligence (AI), and Location Intelligence, concepts in our projects of long-term biodiversity and environmental field monitoring (Fig. 1).

Keywords

standardized data, long term monitoring, biodiversity survey, ecology

Presenting author

Tereza Cristina Giannini

Presented at

TDWG 2021

Author contributions

Conceptualization: TCG, LCT, ALA; Writing - original draft: ALA, TCG, ALA; Review & editing: all authors

Conflicts of interest

None

References

- Henle K, et al. (2013) Priorities for biodiversity monitoring in Europe: A review of supranational policies and a novel scheme for integrative prioritization. *Ecological Indicators* 33: 8-18.
- Kuhl H, et al. (2020) Effective biodiversity monitoring needs a culture of integration. *OneEarth* 3: 462-47.
- Rashed A, et al. (2021) The role of private sector in the implementation of sustainable development goals. *Environment, Development and Sustainability* 23: 2931-3948. <https://doi.org/10.1007/s10668-020-00718-w>
- Shackelford N, et al. (2018) Restoration of Open-Cut Mining in Semi-Arid Systems: A Synthesis of Long-Term Monitoring Data and Implications for Management. *Land Deg. & Dev* 29: 994-1004. <https://doi.org/10.1002/ldr.2746>

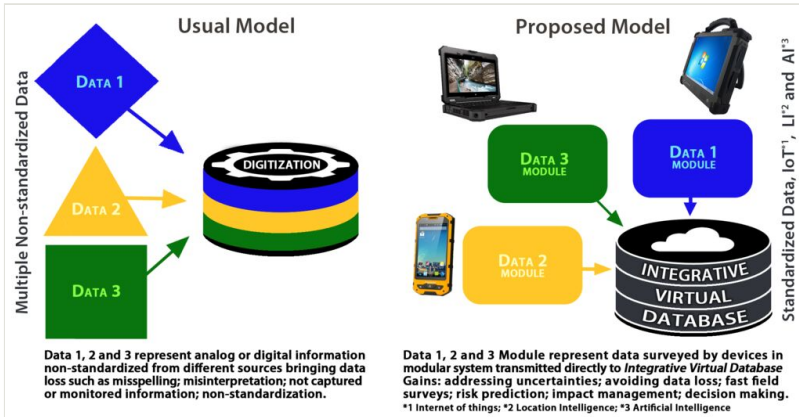


Figure 1.

Infographic: The conventional data collection processes to filling out biodiversity databases can introduce errors and problems such as spelling errors, misinterpretation, typos, and lack of standardization. The use of digital devices to capture and insert standardized information can help to mitigate these problems, bringing efficiency and cost reduction (image made by the authors).