**OR2024 Presentation Submission**

**The 19th International Conference on Open Repositories, June 3-6th 2024, Göteborg, Sweden**

**Understanding DOI attribution in biodiversity repositories: A Brazilian case study**

*Laura Vilela Rodrigues Rezende, Federal University of Goiás (UFG),* *laura\_rezende@ufg.br*

*Geisa Muller de Campos Ribeiro, Federal University of Goiás (UFG),* *geisamuller@ufg.br*

*Maria das Graças Monteiro Castro, Federal University of Goiás (UG),* *gracamcastro@ufg.br*

*Cassia Oliveira, Federal University of Goiás (UFG),* *cassiaoliveira@ufg.br*

*Fabiano Couto Corrêa, Federal University of Rio Grande do Sul (UFRGS),* *fabianocc@gmail.com*

*Carolina Howard Felicissimo, National Network of Research (RNP),* *carolina.felicissimo@rnp.br*

*Keila Elizabeth Macfadem Juarez, National Network of Research (RNP),* *keila.juarez@consultores.rnp.br*

*Clara Baringo Fonseca - National Network of Research (RNP), e-mail:* *clara.fonseca@consultores.rnp.br*

*Yohana Pereira Ditzel - National Network of Research (RNP), e-mail:* *yohana.ditzel@terceiro.rnp.br*

## Abstract

*The data sharing definitions and openness must consider, additionally, issues of institutional interest, national sovereignty, intra- and extra-country asymmetries and of reciprocity, in order to avoid increasing inequalities in the scientific and technology and population access to knowledge. In biodiversity context, Brazil has a huge relevance once the country occupies almost half of South America and is the country with the greatest biodiversity in the world. The repository SiBBr was developed as the Brazilian national repository of data and information on biodiversity, responsible for organizing, indexing, storing and making available data and information about biodiversity and Brazilian ecosystems, providing subsidies for scientific researches and government management related to conservation and sustainable use. This is a Brazilian case study that aims to bring out the challenges involved in understanding DOI attribution in biodiversity repositories*. *How DOI attribution to biodiversity materials* *in SiBBr can work as unique and persistent identifiers, allowing a relevant increase in the citations and visibility of Brazilian biodiversity data. Once the relevant context of SiBBr, serving as the Brazilian national GBIF node, it is mandatory to implement best practices in the SiBBr repository, for instance, including better characterization, identification, location and (re)use of data published.*

## Keywords

*biodiversity data repository; persistent identifier; DOI-Digital Object Identifier;*

## Audience

*biodiversity repository managers, developers, biodiversity data producers, librarians, etc.*

**Proposal**

*The management, sharing and opening of collected or generated research data can be understood as an opportunity to foster collaboration between researchers, between research and innovation institutions and public authorities, the advancement of knowledge and the creation of solutions that meet the demands of society. This is a strategic activity that favors the integrity, quality, reproducibility of research, institutional memory and the reuse of data. The research data management process precedes decisions regarding sharing and opening, and contextual, technical, legal and ethical aspects must be considered. The definitions relating to data sharing and openness must consider, in addition to these aspects, issues of institutional interest, national sovereignty, intra- and extra-country asymmetries and issues of reciprocity, in order to avoid increasing inequalities in the scientific and technology and population access to knowledge.*

*The context of data on biodiversity adds additional relevance when it comes to Brazil, since the country occupies almost half of South America and is the country with one of the greatest biodiversity in the world (Santos et al., 2021). There are more than 52,339 plant species[[1]](#footnote-1) and more than 127,845 animal species[[2]](#footnote-2) known in the country, spread across six terrestrial biomes and three large marine ecosystems. Its different climatic zones favor the formation of biomes (biogeographical zones) and this abundant variety of life is home to more than 20% of the world's total species, found on land and water. After hosting ECO 92 - United Nations Conference on Environment and Development (UNCED) in 1992, held in Rio de Janeiro, the Convention on Biological Diversity – CBD was established, of which Brazil is a signatory. Based on this established framework, a series of commitments have been assumed by the country as a way of operationalizing, mainly, the three pillars of the CBD: the conservation of biological diversity, the sustainable use of biodiversity and the fair and equitable sharing of benefits arising from the use of genetic resources (BRAZIL, 2023).*

*Therefore, in order to monitor changes in biodiversity, it is necessary to collect, document, store and analyze indicators on the spatio-temporal distribution of species, how they interact with each other and with the environment in which they live. Such indicators lack a framework of reliable data from expeditions, biological collections, sensors, scientific publications, among others.*

*SiBBr[[3]](#footnote-3) was developed as the Brazilian national repository of data and information on biodiversity, responsible for organizing, indexing, storing and making available data and information about biodiversity and Brazilian ecosystems, providing subsidies for government management related to conservation and sustainable use (SiBBr, 2023). It acts as the Brazilian node of GBIF[[4]](#footnote-4) (Global Biodiversity Information Facility), a multilateral initiative of approximately 60 participating countries. To get an idea of ​​the importance of this network of systems on biodiversity, recently, a study carried out by Deloitte demonstrated, from an economic point of view, the importance of data mediated by GBIF: for every 1 euro invested in its infrastructure, 3€ are reversed in benefits for platform users and up to €12 for society (DELOITTE, 2023, p.6).*

*SiBBr adopts international standards and protocols for sharing data and information and uses the Australian open source Living Atlas - ALA infrastructure[[5]](#footnote-5), with several modules that can be reused by other organizations. The data available on the platform comes from national research institutions, public or private, research projects and programs and thematic networks (data providers) (SiBBr, 2023).*

*Once the relevant context of SiBBr, serving as the Brazilian national GBIF node, it is mandatory to implement best practices in the SiBBr repository, for instance, including better characterization, identification, location and (re)use of data published. This study aims to bring out the challenges involved in understanding DOI attribution in biodiversity repositories: A Brazilian case study*. *How DOI attribution to biodiversity materials* *in SiBBr* can work as *unique and persistent identifiers, allowing a relevant increase in the citations and visibility of Brazilian biodiversity data.*

*Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications. The Data Citation Principles, have since 2014 been endorsed by 125 institutions including Publishers and funders, cover purpose, function and attributes of citations. These principles recognize the dual necessity of creating citation practices that are both human understandable and machine-actionable. These citation principles are not comprehensive recommendations for data stewardship. And, as practices vary across communities and technologies will evolve over time, we do not include recommendations for specific implementations, but encourage communities to develop practices and tools that embody these principles.*

*Data citations should facilitate identification of, access to, and verification of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verifying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited.*

*Thus, based on the case of data published in the GBIF repository network, several particularities deserve to be presented, aiming for a better understanding of the context of identifying research objects and citations in this knowledge area of biodiversity worldwide. Below are descriptions of the four classes of datasets, identified by DOIs, currently proposed and supported by GBIF:*

* *Resources metadata*

*At its simplest level, GBIF.org allows institutions to create datasets, describing undigitized resources. All three other dataset classes include this basic information, but this ‘metadata-only’ class offers researchers a valuable tool for discovering and learning about evidence not yet available online.*

* *Checklist data*

*Datasets can also provide a catalog or list of named organisms or taxa. While they may include additional details like local species names or specimen citations, these ‘checklists’ typically categorize information along taxonomic, geographic, and thematic lines, or some combination of the three.*

* *Occurrence data*

*Other datasets published through GBIF.org have sufficiently consistent detail to contribute information about the location of individual organisms in time and space—that is, they offer evidence of the occurrence of a species (or other taxon) at a particular place on a specified date. Occurrence datasets make up the core of data published through GBIF.org, and examples can range from specimens and fossils in natural history collections, observations by field researchers and citizen scientists, and data gathered from camera traps or remote-sensing satellites.*

* *Sampling-event data*

*Datasets sometimes provide greater detail, not only offering evidence that a species occurred at a given location and date, but also making it possible to assess community composition for broader taxonomic groups or even the abundance of species at multiple times and places. These quantitative or sampling-event datasets, typically derive from standard protocols for measuring and monitoring biodiversity like vegetation transects, bird censuses and freshwater or marine sampling.*

*In addition for defining the types of datasets available in GBIF and which receive DOIs when published, it is necessary to present, according to what is contained in the repository, which types of digital objects can be persistently identified by DOIs:*

* *Datasets presented from GBIF network;*
* *Data downloaded from GBIF.org;*
* *Research articles and reports published by scientific journals, agencies and NGOs;*
* *Materials deposited in a general-use repository.*

*In the cases listed above, DOIs attribution results from the search and download functionalities made in GBIF and this fact stands out[[6]](#footnote-6).*

*This is a particularity observed in the area of biodiversity due to the frequent need of work with a large amount of data, which often be unfeasible to be cited from unitary repositories in a scientific article. The source of information comes from a diversity of repositories, where data is spread out. The clippings present in the result of searches, from the download page in GBIF network, make up a relevant digital object in the context of scientific work on biodiversity.*

*This work has been carrying out a bibliographic and documentary survey on the attribution of DOIs in the context of biodiversity in Brazil, aiming to define the best strategy for offering this resource in SiBBr. It is also planned that the aforementioned repository will receive records from traditional knowledge of local communities and indigenous peoples associated with species registered in Brazil. This brings out not only the need of properly applying the FAIR principles with DOI attribution, but, mainly, the need of applying tk labels[[7]](#footnote-7), for instance, implementing the CARE principles related to collective benefit, authority to control, responsibility and ethics (Carrol et al.2021).*

*In the context of Open Repositories 2024 Conference, this study contemplates the three main thematics once prioritizes the improvement of research transparency, by promoting persistent identifiers in datasets, elevates underrepresented communities and sustainability, once SiBBr will provide records from traditional knowledge of local communities and indigenous peoples associated with species registered in Brazil according to the protocols previously established.*

## References

*Ministério do Meio Ambiente (BR). Biodiversidade [Internet] 2023. Available from:* [*https://antigo.mma.gov.br/temas/biodiversidade.html*](https://antigo.mma.gov.br/temas/biodiversidade.html)

 *CARROLL S R, Herczog E, Hudson M, Russell K, Stall S. Operationalizing the CARE and FAIR Principles for Indigenous data futures. Scientific Data. 2021; 8(108): 1-6, vol. 8, no 1, 108 p. 2021. Available from:* [*https://doi.org/10.1038/s41597-021-00892-0*](https://doi.org/10.1038/s41597-021-00892-0)

*Data Citation Synthesis Group. Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014* [*https://doi.org/10.25490/a97f-egyk*](https://doi.org/10.25490/a97f-egyk)

*Delloite Access Economics. GBIF Secretariat. Economic valuation and assessment of the impact of the GBIF network; 2023. Available from:* [*https://www.deloitte.com/content/dam/assets-zone1/au/en/docs/services/economics/deloitte-economics-global-biodiversity-information-facility-260623.pdf*](https://www.deloitte.com/content/dam/assets-zone1/au/en/docs/services/economics/deloitte-economics-global-biodiversity-information-facility-260623.pdf)

 *Sistema de Informação sobre a Biodiversidade Brasileira. O que é o SiBBr [Internet]. 2023. Available from:* [*https://sibbr.gov.br/page/o-que-sibbr.html*](https://sibbr.gov.br/page/o-que-sibbr.html)*.*

*Santos, M. E. d., Silva, M. M., Pereira, A. F., & Silva, T. T. C. d. 2021. Calendário de sazonalidade de hortifrutícolas não convencionais ou negligenciadas. DEMETRA: Alimentação, Nutrição &Amp; Saúde. Available from:* [*https://doi.org/10.12957/demetra.2021.48506*](https://doi.org/10.12957/demetra.2021.48506)

1. [*https://floradobrasil.jbrj.gov.br/reflora/listaBrasil/PrincipalUC/PrincipalUC.do;jsessionid=2CCECDC0C6A6D6D7B215F968D693B89A#CondicaoTaxonCP*](https://floradobrasil.jbrj.gov.br/reflora/listaBrasil/PrincipalUC/PrincipalUC.do;jsessionid=2CCECDC0C6A6D6D7B215F968D693B89A#CondicaoTaxonCP) [↑](#footnote-ref-1)
2. [*http://fauna.jbrj.gov.br/fauna/listaBrasil/ConsultaPublicaUC/ResultadoDaConsultaNovaConsulta.do*](http://fauna.jbrj.gov.br/fauna/listaBrasil/ConsultaPublicaUC/ResultadoDaConsultaNovaConsulta.do) [↑](#footnote-ref-2)
3. [*https://www.sibbr.gov.br/*](https://www.sibbr.gov.br/) [↑](#footnote-ref-3)
4. [*https://www.gbif.org/*](https://www.gbif.org/) [↑](#footnote-ref-4)
5. [*https://living-atlases.gbif.org/*](https://living-atlases.gbif.org/) [↑](#footnote-ref-5)
6. *For instance, while querying about “Paubrasilia” genus the DOI: 10.15468/dl.usxz5t is generated and its persistent url is given for citation in publications*

*“GBIF present PLEASE USE THIS CITATION IN PUBLICATIONS*

 *GBIF Occurrence Download https://doi.org/10.15468/dl.usxz5t”* [↑](#footnote-ref-6)
7. [*https://localcontexts.org/labels/traditional-knowledge-labels/*](https://localcontexts.org/labels/traditional-knowledge-labels/) [↑](#footnote-ref-7)