BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 9: e177966
doi: 10.3897/biss.9.177966

OPEN ACCESS

Conference Abstract

Everywhere Everyone Everything All at Once: Integrating Data Infrastructures and Analysis Workflows for the Upscaling to Global Genetic Monitoring

Jutta Buschbom[‡], Elizabeth R. Ellwood[§], Sharif Islam[|,¶], Eva Häffner[#], Alexandra Pavlova[¤]

‡ Statistical Genetics, Ahrensburg, Germany
§ iDigBio, Gainesville, Florida, United States of America
| Naturalis Biodiversity Center, Leiden, Netherlands
¶ DiSSCo, Leiden, Netherlands
# Botanischer Garten und Botanisches Museum Berlin, Freie Universität Berlin, Berlin, Germany
¤ Wildlife Genetic Management Hub, School of Biological Sciences, Monash University, Melbourne, Australia

Corresponding author: Jutta Buschbom (jutta.buschbom@statistical-genetics.de)

Received: 13 Nov 2025 | Published: 19 Nov 2025

Citation: Buschbom J, Ellwood ER, Islam S, Häffner E, Pavlova A (2025) Everywhere Everyone Everything All at Once: Integrating Data Infrastructures and Analysis Workflows for the Upscaling to Global Genetic Monitoring. Biodiversity Information Science and Standards 9: e177966. https://doi.org/10.3897/biss.9.177966

Abstract

Effective decision-making on biodiversity restoration would greatly benefit from baseline data on intraspecific genetic diversity, the ability to integrate it across species for each location, and efficient systems for monitoring changes of genetic diversity in response to management interventions and environmental dynamics.

This requires large sets of well-curated information-rich FAIR (Findable, Accessible, Interoperable, Reusable) Digital Objects (FDOs; Schultes and Wittenburg 2019), implemented as, for instance, Digital Extended Specimens (DES), which are representing digitized field samples, their derived genomic data and associated information (Hardisty et al. 2022). These use-case-driven sets of structured (meta)data from many providers need to be merged, modified and further extended on demand. Existing workflows and work environments have to be redesigned to accelerate this process and achieve

This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

seamless integration to be able to scale to the needs of efficient and effective worldwide monitoring under the United Nations Kunming-Montreal Global Biodiversity Framework.

Contributing to data and infrastructure development processes in support of global monitoring is also the global effort to generate high-quality reference genomes. The Earth BioGenome communities, including the European Reference Genome Atlas (ERGA) and Biodiversity Genomics Europe (BGE) communities, drive the development of standardization and harmonization for well-designed sampling and comprehensive FAIR and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) (meta)data (Buzan et al. 2025). Their goal is reproducible analytical pipelines that can be assembled on the fly for implementing sophisticated statistical approaches and algorithms.

Such research-focused pipelines prepare the development and application of globally adopted workflow templates for the calculation of Essential Biodiversity Variables (EBVs). These templates are under development within the Group on Earth Observations Biodiversity Observation Network (GEO BON) (Lumbierres et al. 2025). Their output can be used for globally aligned and interpretable planning, monitoring, reporting and reviewing (see CBD/COP/16/L.33).

Achieving global agreement on a (small) set of jointly used interoperable vocabularies and ontologies for (meta)data and machine-actionable operations is a communication, community-capacity development, and negotiation process that requires significant resources, engagement across sociocultural groups and geographies, patience and time. Ongoing processes towards these goals continue to be organized and promoted by, e.g., Biodiversity Information Standards (TDWG), the Global Biodiversity Information Facility (GBIF), GEO BON, and the Ocean Biodiversity Information System (OBIS), as well as large continental networks.

We propose to bridge the gap before standardization and harmonization are in place and thereby facilitate and accelerate such efforts. Taking a pragmatic approach, our objective is to contribute a lightweight "pocket" dataspace that provides interoperability and data governance in connection with a digital platform for global genetic monitoring (Fig. 1).

The pocket dataspace would allow existing platforms and tools to be connected easily through community-provided mappings between workflow element-specific formats, terms, data and operations stored in an open repository. This approach would enable users to take advantage of the core strengths of existing software products and the expertise of their associated communities, while quickly sharing data and their work between specialized solutions. At the same time, data would be FAIRified and CAREd-for, promoting attribution, transparency and responsibility. The functions of the pocket dataspace can be prerequisites for a transition to machine-actionable operations usable to agentic AI. As a general-purpose interlinking and translation component, the pocket dataspace aims to be the missing link between distributed, federated, non-standard-compliant and undocumented data, governance regimes, provenance logs and software output, and the need for transparent, well-governed and versatile conservation applications.

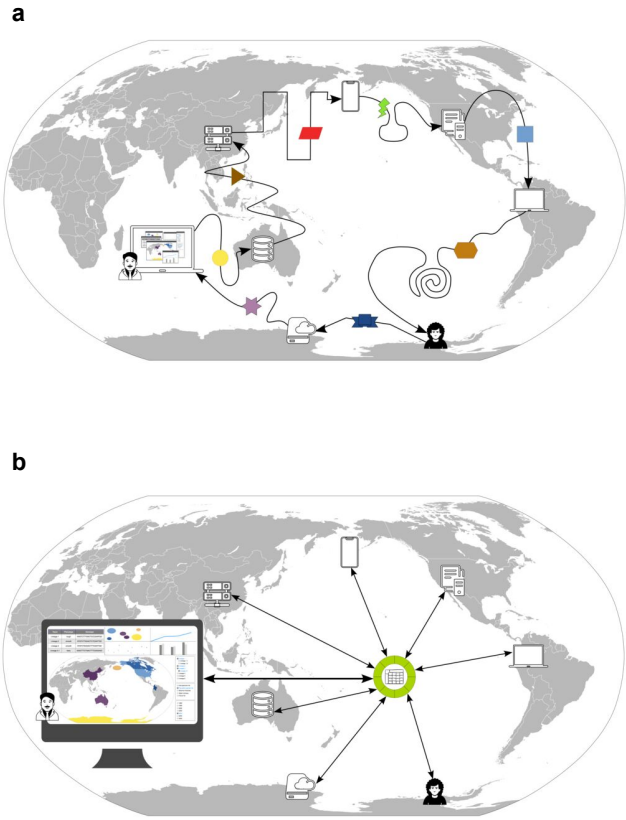

Figure 1.

Conceptualization of mobilizing existing data and software resources through a "pocket" dataspace in combination with a global monitoring platform for genetic diversity. The upper graphic depicts the current state of manual workflows, the lower graphic shows the effect of the proposed pocket dataspace. BlankMap-World-162E.svg by Lokal_Profil on Wikipedia, public domain; icons and mockups from uxwing and drawio.

One of these conservation applications will be the proposed platform for monitoring global genetic diversity. The platform will aggregate and visualize externally-linked population-genetic data and summary metrics that are the results of analysis pipelines enabled by, e.g., the pocket dataspace. Its aim is to provide visualization and support dataset and analysis management for local to global conservation efforts. It would store uploaded or linked genetic diversity metrics, perform selected automated analyses for continuously updated genetic diversity measures, as well as provide a starting point for user-designed analyses. The objective of our initial use case is to analyze three basic measures of population-genetic diversity based on genome-wide sequencing data as a first step towards operationalizing genetic monitoring at scale.

Together, the pocket dataspace and monitoring platform for genetic diversity data have the goal to support a digital ecosystem that is foremost flexible, requiring low investments by users, and be able to quickly integrate both inter- and transdisciplinary data as well as existing powerful platforms and well-tested analytical pipelines and functionality.

Keywords

biodiversity restoration, cooperative platform, flexible workflows, dataspace, FAIR, CARE, agentic AI

Presenting author

Jutta Buschbom

Presented at

Living Data 2025

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Buzan E, de Guttry C, Bortoluzzi C, et al. (2025) Biodiversity Genomics Research Practices Require Harmonising to Meet Stakeholder Needs in Conservation. Molecular Ecology https://doi.org/10.1111/mec.70001
- Hardisty AR, Ellwood ER, Nelson G, et al. (2022) Digital Extended Specimens: Enabling an Extensible Network of Biodiversity Data Records as Integrated Digital Objects on the Internet. BioScience 72 (10): 978-987. https://doi.org/10.1093/biosci/biac060
- Lumbierres M, Milanović M, Beja P, et al. (2025) Towards implementing workflows for essential biodiversity variables at a European scale. Global Ecology and Conservation 62 https://doi.org/10.1016/j.gecco.2025.e03699
- Schultes E, Wittenburg P (2019) FAIR Principles and Digital Objects: Accelerating Convergence on a Data Infrastructure. Communications in Computer and Information Science 1003: 3-16. https://doi.org/10.1007/978-3-030-23584-0_1