# Summary Report
Ecological Data Listening Session | July 12, 2024

The Ecological data-focused listening session of the BIOFAIR Data Network Project held on July 12, 2024 was led by BIOFAIR Data Network Steering Committee members Matthew Sheik (Denver Botanic Gardens), Mike Webster (Cornell Lab of Ornithology), Libby Ellwood (iDigBio), Nico Franz (University of Kansas Biodiversity Institute), and Jyotsna Pandey (American Institute of Biological Sciences), in collaboration with Key Domain Representatives Gregory Maurer (Jornada Basin Long-Term Ecological Research site, Environmental Data Initiative) and Stephanie Parker (NEON: National Ecological Observatory Network).

Among the 41 session participants were representatives from Archbold Biological Station, Ecological Society of America, Smithsonian Institution, University of Arizona, National Ecological Observatory Network (NEON), USA National Phenology Network, Environmental Data Initiative, Long Term Ecological Research (LTER) Network, Polyneme LLC, Perot Museum of Nature and Science, Berkeley Lab, National Microbiome Data Collaborative (NMDC), National Oceanic and Atmospheric Administration (NOAA), University of Texas at Austin, United States Department of Agriculture (USDA), Atlantic International Research (AIR) Centre, Rauthiflor LLC, Los Alamos National Laboratory, Global Biodiversity Information Facility (GBIF), iDigBio, Denver Botanic Gardens, Cornell University, University of California - San Diego, Pacific Northwest National Lab, Michigan Natural Features Inventory, National Science Foundation (NSF), Lawrence Berkeley National Laboratory (LBNL), Carnegie Science, U.S. Geological Survey, Ocean Biodiversity Information System (OBIS) - USA, The University of Texas Marine Science Institute, and Environmental Molecular Sciences Laboratory (EMSL) - Pacific Northwest Laboratory.

## Summary

The BIOFAIR Ecological Data Listening Session brought together ecological data experts from various organizations to discuss issues related to biodiversity data management and integration. The session opened with presentations from representatives of the Long Term Ecological Research Network and the National Ecological Observatory Network, which served as a baseline for discussion about the complexities of ecological data, as well as integration of ecological data with other types of data. In full- and small-group rooms, participants discussed the importance of data integration, the challenges of cross-scale and cross-platform integration, and the potential for increased investment in standards, ontology development, and the adoption of translation tools to operate among services.

**Key Topics Discussed**

**Introduction to the BIOFAIR Data Network Project and the Ecological Data Listening Session**

The session began with a brief introduction by Matthew Sheik to BCoN and the NSF-funded BIOFAIR Data Network project (Award No. 2303588). The BCoN Steering Committee is part of a broader community representing various traditional museum groups, botanic gardens, paleontologists, culture collections, zoos, and other types of biodiversity collections. BCoN aims to promote the integration, use, and impact of biodiversity data and collections. The group has promoted the development of an Extended Specimen Network as a unifying goal for biodiversity collections over the next decade. The Ecological Data Listening Session was part of a series of six domain-focused sessions for catalyzing collaborations towards building a Findable, Accessible, Interoperable, and Reusable (FAIR) data network. The listening sessions will be followed by an interdisciplinary virtual workshop in early 2025 to delve deeper into discussions about a collaborative path forward. The goal of this workshop will be to bring together an expansive, non-conventional configuration of ecological data aggregators and users to work together to develop recommendations and create a roadmap that outlines the gaps, needs, and actions needed to build a FAIR, open, integrated biological and ecological data network.

**Data Integration and Cross-Scale Challenges**

As a full-group, participants discussed data integration, emphasizing its importance for interoperability and answering a diversity of research questions. There was agreement that data integration doesn't require storing all data in one place, but rather the ability to transform data and combine it with the right metadata for integrated analysis. Discussion included the challenges of cross-scale integration, particularly with remote sensing data and the recent expansion of occurrence data into sampling event data. Attendees identified various ecological data stakeholders that should be part of the effort to build an expanded data network, with suggestions being captured in a [brainstorming document](#).

**LTER Network and NEON Discussion**

Gregory Maurer, the Information Manager for the Jornada Basin Long Term Ecological Research (LTER) site, discussed the LTER Network, its research areas, and the recent addition of a specimen management plan requirement for site renewals. Maurer also emphasized the importance of data integration (usually termed "synthesis" in LTER), reuse, and the network's role in pioneering data publication and FAIR data standards. Next, Stephanie Parker, an aquatic ecologist working in the Aquatic Observation System team at NEON, presented on NEON, a continental-scale project with 81 sites across the US and its territories, designed to observe and forecast the impacts of climate and environmental change. Parker highlighted NEON's plans to further partnerships and make its data available on external data repositories to reach a wider audience.

**Discussion on Ecological Data Integration Challenges, Resource Trade-Offs, Solutions**

Following the presentations, participants were divided into groups for discussions on challenges and areas to make progress in integrating ecological data with other types of data, considering resource trade-offs in the immediate, short-, and long-term. After the breakout sessions, the groups reconvened to share their findings. Participants discussed numerous challenges and potential solutions for data integration, with outcomes focused on the need for common language, flexibility in data models, and the importance of dedicated managers for species data. The idea of developing translators across different repositories was proposed to manage data heterogeneity and connect networks of ecological data, and the importance of increased investment in standards and ontology development, data tracking, and the need for sustained funding for sample and data management was emphasized. Another issue raised was the challenge of getting researchers to collect data in a way that can be integrated with other datasets. The goals of research projects that require specimen collection might not necessarily overlap with the goals of data integration. DataOne (https://www.dataone.org/), which federates existing ecological data repositories, was suggested as a potential model to inform future integration. The general idea would be that various datasets are searchable from one place, while being retained in many different repositories.

**Data Standards, Collaborations, and Emerging Data Types**

Participants discussed the development and adoption of data standards to improve the accessibility and communication of biodiversity and ocean data. They also explored potential collaborations, such as with the Environmental Data Initiative and the LTER Network, and the importance of standardizing metadata and harmonizing data for interoperability. Additionally, the group discussed hypothesis driven data integration rather than 'data integration for data integration's sake.' The discussion also touched on emerging data types such as bioacoustic data and camera trap data, the challenges of integrating these into existing systems, and the potential of leveraging existing data repositories. Participants emphasized the importance of sample types, participatory science, and the need for standardized protocols and vocabularies when combining disparate datasets.

# Recommendations

- Ecological data communities, including NEON and LTER, in conjunction with the biodiversity informatics community, need to continue developing and adopting data and research metadata standards, which will enable harmonization of data across networks and interoperability of ecological data repositories with biodiversity informatics systems.
- Community and participatory science data, and the complexities of it, should be included in discussions about ecological data integration.
- The Humboldt Core extension should be considered to improve ecological data integration and to be inclusive of, for e.g., acoustic monitoring data including freshwater, camera trap data, animal movement data, etc.

- Data repositories should consider implementing translation layers among different data standards and formats.
- Research communities should consider a shift in priorities towards hiring data managers and include data management skills in career advancement. Likewise, educational institutions should prioritize data management and standards training for early career scientists.
- Success stories of data integration efforts should be highlighted and promoted as examples.
- A roadmap or concept map of data integration efforts, including users, contributors, repositories, and data aggregators, should be developed to guide future data integration.